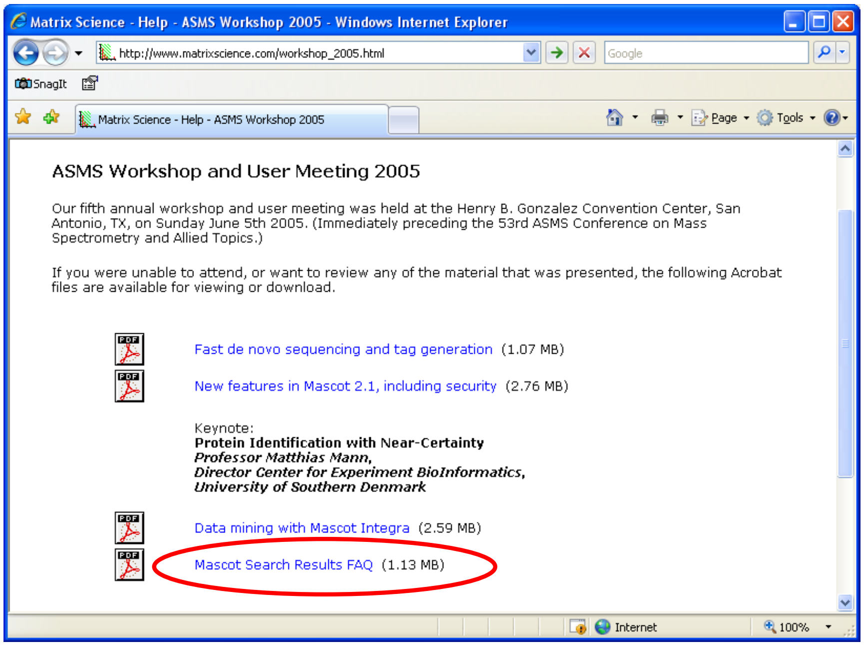
Mascot Search Results FAQ

MASCOT

{MATRIX}
{SCIENCE}



ASMS Workshop and User Meeting 2005

Our fifth annual workshop and user meeting was held at the Henry B. Gonzalez Convention Center, San Antonio, TX, on Sunday June 5th 2005. (Immediately preceding the 53rd ASMS Conference on Mass Spectrometry and Allied Topics.)

If you were unable to attend, or want to review any of the material that was presented, the following Acrobat files are available for viewing or download.

Fast de novo sequencing and tag generation (1.07 MB)

New features in Mascot 2.1, including security (2.76 MB)

Keynote:
Protein Identification with Near-Certainty
Professor Matthias Mann,
Director Center for Experiment BioInformatics,
University of Southern Denmark

Data mining with Mascot Integra (2.59 MB)

Mascot Search Results FAQ (1.13 MB)

We had a presentation with this same title at our 2005 user meeting. So much has changed in the last 6 years that it seemed like a good idea to re-visit the topic. Just about everything in the earlier talk is still relevant to the current version of Mascot, but three of the original answers would benefit from being updated

Why are some peptide matches shown in red or bold face?

Red indicates the top scoring peptide match for this spectrum

- Not necessarily a significant match

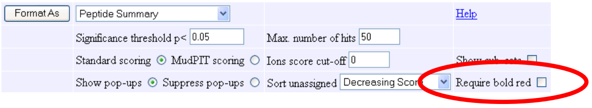
Bold indicates the first time any match to this spectrum has appeared in the report

- If a protein hit doesn't have any bold red matches, all the assigned spectra have better scoring matches elsewhere or the same matches have already appeared in the report, assigned to higher scoring protein(s)

This is still true for Peptide Summary and Select Summary reports

We thought his was probably the most frequently asked question. The original answer, given here, is still correct for the Peptide Summary and Select Summary reports.

Why are some peptide matches shown in red or bold face?

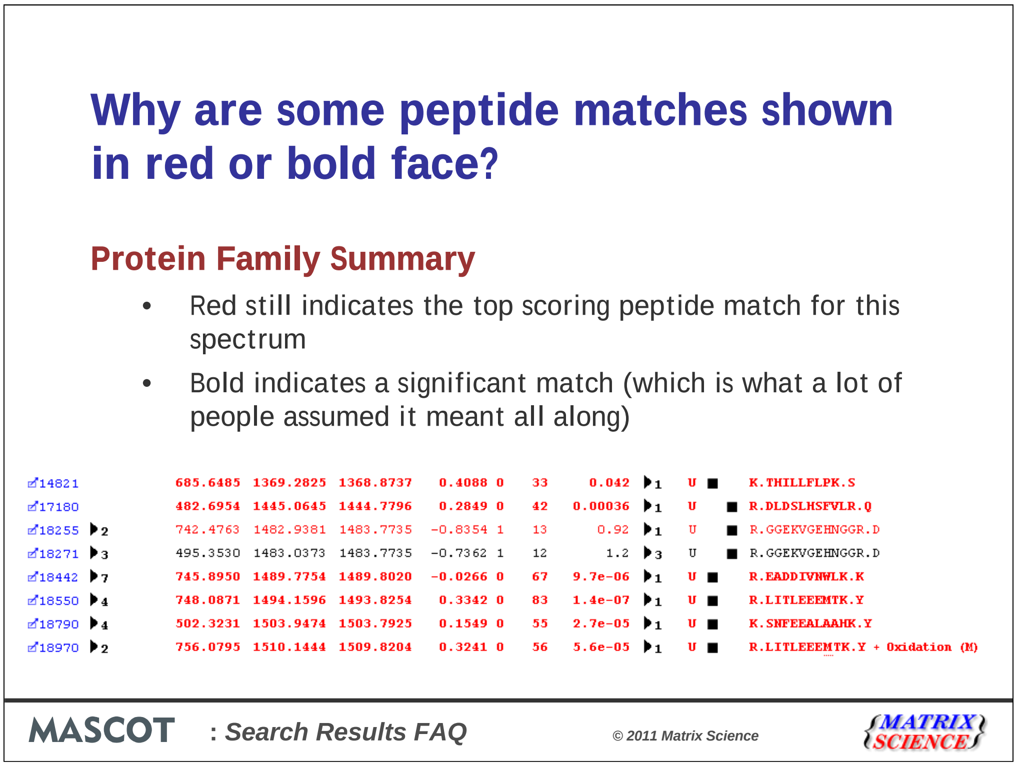


Format As | Peptide Summary | Help

Significance threshold p< 0.05
Max. number of hits 50
Standard scoring ● MudPIT scoring ○
Ions score cut-off 0
Show sub-sets ☐
Show pop-ups ● Suppress pop-ups ○
Sort unassigned Decreasing Score
Require bold red ☐

20: Q3UWB9	34: Q91WH2	51: Q3UEP4
GAAVALNIR		
	GAAVTLNIR	GAAVTLNIR
	AEMWLIR	AEMWLIR
IILDELVQR	IILDELVQR	
FSPGYQIEK	FSPGYQIEK	
DNLENFFIK	DNLENFFIK	
FETFPTSVSK	FETFPTSVSK	
FVDVWTYEMPR	FVDVWTYEMPR	
		WTYEVPR
		IILDELK
		TPATLGPNTR
		FSPGYYLEK
WLPQNDLLGHPK	WLPQNDLLGHPK	WLPQNDLLGHPK
ANAIAWALAQIPQK		ANAIAWALAQIPQK
GHEVTVLRPSAYYVLDPK	GHEVTVLRPSAYYVLDPK	

By dropping hits that have no bold red matches, we can largely eliminate two classes of false proteins. First, proteins that would be sub-sets of other, higher scoring proteins except for the presence of one or more low scoring, random peptide matches. Second, intersection proteins, that are sub-sets of two or more protein. However, ‘require bold red’ doesn’t always work. Here, for example, hit 34 is an intersection protein. All its matches are found in hits 20 and 51. But, because of the order in which they appear, they all have bold red matches.



Why are some peptide matches shown in red or bold face?

Protein Family Summary

- Red still indicates the top scoring peptide match for this spectrum
- Bold indicates a significant match (which is what a lot of people assumed it meant all along)

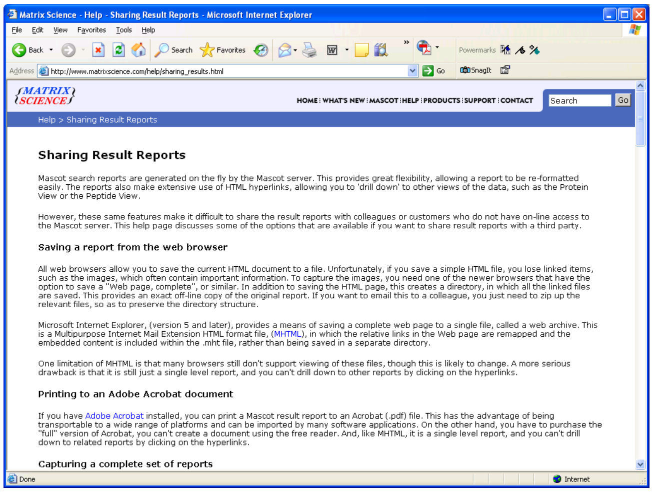
```
14821        685.6485 1369.2825 1368.8737   0.4088 0   33     0.042 ▶1  U ■    K.THILLFLPK.S
17180        482.6954 1445.0645 1444.7796   0.2849 0   42   0.00036 ▶1  U    ■ R.DLDSLHSFVLR.Q
18255 ▶2     742.4763 1482.9381 1483.7735  -0.8354 1   13      0.92 ▶1  U    ■ R.GGEKVGEHNGGR.D
18271 ▶3     495.3530 1483.0373 1483.7735  -0.7362 1   12       1.2 ▶3  U    ■ R.GGEKVGEHNGGR.D
18442 ▶7     745.8950 1489.7754 1489.8020  -0.0266 0   67   9.7e-06 ▶1  U ■    R.EADDIVNWLK.K
18550 ▶4     748.0871 1494.1596 1493.8254   0.3342 0   83   1.4e-07 ▶1  U ■    R.LITLEEEMTK.Y
18790 ▶4     502.3231 1503.9474 1503.7925   0.1549 0   55   2.7e-05 ▶1  U ■    K.SNFEEALAAHK.Y
18970 ▶2     756.0795 1510.1444 1509.8204   0.3241 0   56   5.6e-05 ▶1  U ■    R.LITLEEEMTK.Y + Oxidation (M)
```

MASCOT : *Search Results FAQ*

The Protein Family Summary, introduced in Mascot 2.3, was developed to present a more rigorous minimal list of proteins, without the complication of worrying about bold red. We described it in details at last year’s meeting, in a talk that can be found on our web site. Proteins are grouped into families on the basis of shared peptide matches, and any intersection proteins identified and eliminated. Low scoring matches cannot interfere because grouping is based on significant matches only.

This means that the earlier use of bold is no longer required. In the new report, bold is used for matches that are significant, which is what a lot of people assumed it meant all along.

How can I send a result report to a colleague?



Sharing Result Reports

Mascot search reports are generated on the fly by the Mascot server. This provides great flexibility, allowing a report to be re-formatted easily. The reports also make extensive use of HTML hyperlinks, allowing you to 'drill down' to other views of the data, such as the Protein View or the Peptide View.

However, these same features make it difficult to share the result reports with colleagues or customers who do not have on-line access to the Mascot server. This help page discusses some of the options that are available if you want to share result reports with a third party.

Saving a report from the web browser

All web browsers allow you to save the current HTML document to a file. Unfortunately, if you save a simple HTML file, you lose linked items, such as the images, which often contain important information. To capture the images, you need one of the newer browsers that have the option to save a "Web page, complete", or similar. In addition to saving the HTML page, this creates a directory, in which all the linked files are saved. This provides an exact off-line copy of the original report. If you want to email this to a colleague, you just need to zip up the relevant files, so as to preserve the directory structure.

Microsoft Internet Explorer, (version 5 and later), provides a means of saving a complete web page to a single file, called a web archive. This is a Multipurpose Internet Mail Extension HTML format file, (MHTML), in which the relative links in the Web page are remapped and the embedded content is included within the .mht file, rather than being saved in a separate directory.

One limitation of MHTML is that many browsers still don't support viewing of these files, though this is likely to change. A more serious drawback is that it is still just a single level report, and you can't drill down to other reports by clicking on the hyperlinks.

Printing to an Adobe Acrobat document

If you have Adobe Acrobat installed, you can print a Mascot result report to an Acrobat (.pdf) file. This has the advantage of being transportable to a wide range of platforms and can be imported by many software applications. On the other hand, you have to purchase the "full" version of Acrobat, you can't create a document using the free reader. And, like MHTML, it is a single level report, and you can't drill down to related reports by clicking on the hyperlinks.

Capturing a complete set of reports

Another FAQ dealt with sending search result reports to colleagues or customers. The answers discussed in 2005 may still be applicable in some cases, and are described on this help page, but there are now two additional options.

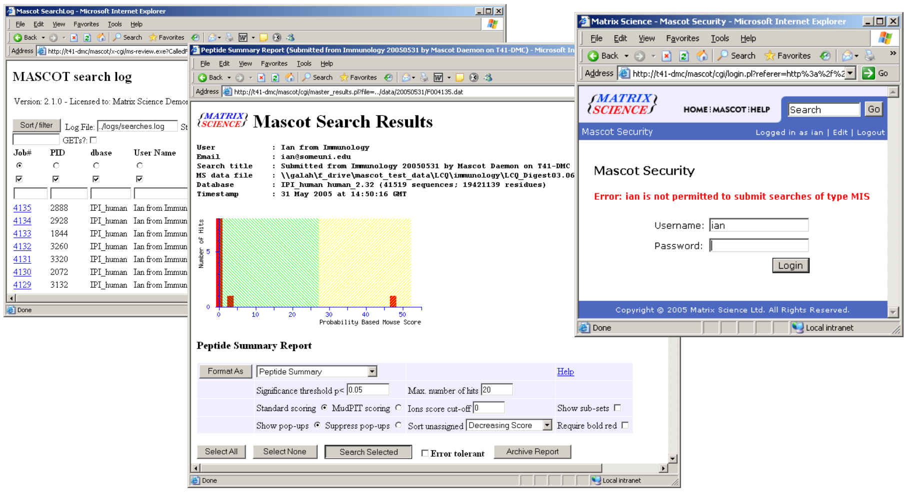
How can I send a result report to a colleague?

Mascot Security (new in Mascot 2.1) enables you to configure users who can:

- View their own reports
- Not see reports belonging to users in other groups
- Not submit searches

Often, the most versatile option is to allow the colleague to connect to the Mascot server, but with restricted rights.

How can I send a result report to a colleague?



MASCOT search log

Version: 2.1.0 - Licensed to: Matrix Science Demo

Sort / filter Log File: ../logs/searches.log

GETs?: ☐

Job#	PID	dbase	User Name
●	○	○	○
☑	☑	☑	☑
4135	2888	IPI_human	Ian from Immun
4134	2928	IPI_human	Ian from Immun
4133	1844	IPI_human	Ian from Immun
4132	3260	IPI_human	Ian from Immun
4131	3320	IPI_human	Ian from Immun
4130	2072	IPI_human	Ian from Immun
4129	3132	IPI_human	Ian from Immun

Mascot Search Results

```
User          : Ian from Immunology
Email         : ian@someuni.edu
Search title  : Submitted from Immunology 20050531 by Mascot Daemon on T41-DMC
MS data file  : \\galah\f_drive\mascot_test_data\LCQ\immunology\LCQ_Digest03.06
Database      : IPI_human human_2.32 (41519 sequences; 19421139 residues)
Timestamp     : 31 May 2005 at 14:50:16 GMT
```

Peptide Summary Report

Format As: Peptide Summary — Help

Significance threshold p< 0.05 — Max. number of hits 20

Standard scoring ● MudPIT scoring ○ — Ions score cut-off 0 — Show sub-sets ☐

Show pop-ups ● Suppress pop-ups ○ — Sort unassigned Decreasing Score — Require bold red ☐

Select All — Select None — Search Selected — ☐ Error tolerant — Archive Report

Mascot Security — Logged in as ian | Edit | Logout

Mascot Security

Error: ian is not permitted to submit searches of type MIS

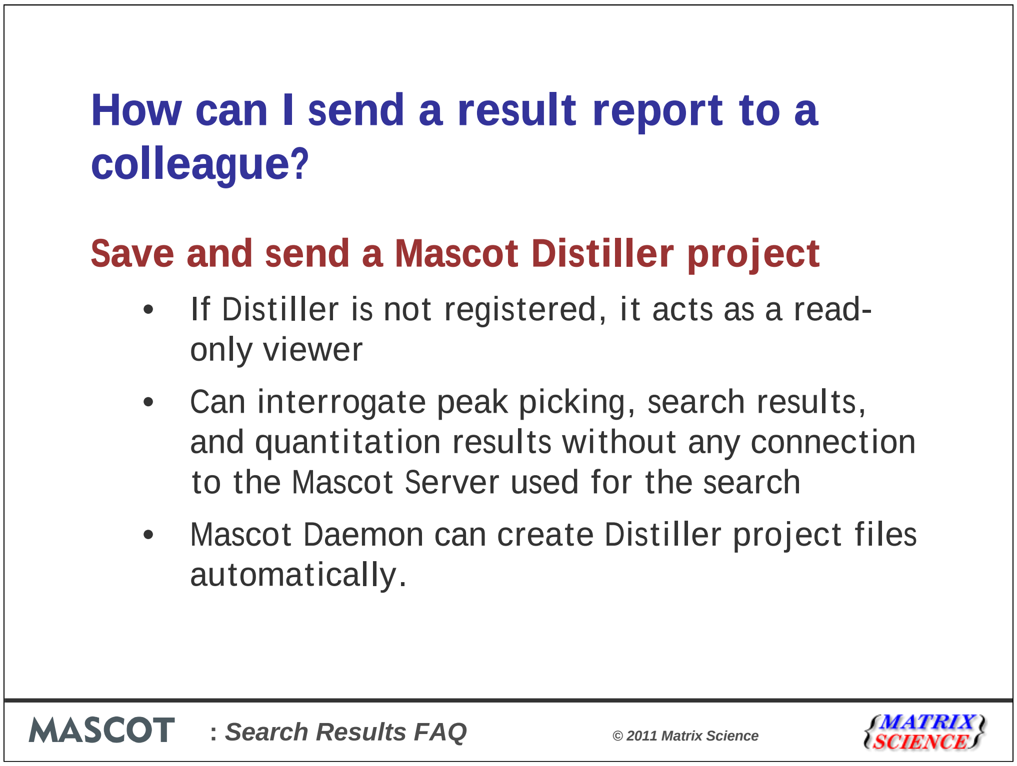
Username: ian

Password:

Login

Copyright © 2005 Matrix Science Ltd. All Rights Reserved.

When such a user goes to the search log, they will see their own search results, but not other people’s. Even if they have a link to someone else’s results, they cannot open them. If they try to do a repeat search, or a new search, they are requested to log in as someone with the rights to run searches



How can I send a result report to a colleague?

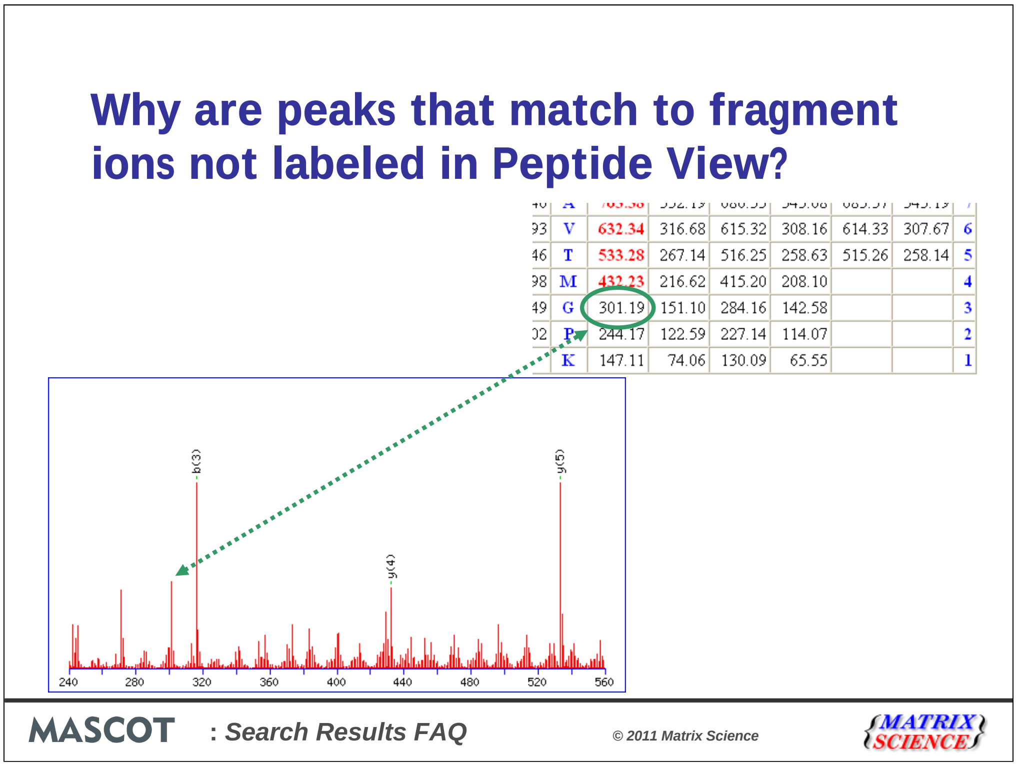
Save and send a Mascot Distiller project

- If Distiller is not registered, it acts as a read-only viewer
- Can interrogate peak picking, search results, and quantitation results without any connection to the Mascot Server used for the search
- Mascot Daemon can create Distiller project files automatically.

MASCOT : *Search Results FAQ*

MATRIX SCIENCE

If this isn’t appropriate, then why not save the search results to a Mascot Distiller project and send the project file to your colleague? If Distiller is not registered, it is a free, read-only viewer for existing projects. As well as search results, you can see the details of peak picking and any saved quantitation results. If you have the Daemon Toolbox option for Distiller, you can batch process files and automatically save Distiller projects.



Why are peaks that match to fragment ions not labeled in Peptide View?

[illegible]	A	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	7
93	V	632.34	316.68	615.32	308.16	614.33	307.67	6
46	T	533.28	267.14	516.25	258.63	515.26	258.14	5
98	M	432.23	216.62	415.20	208.10			4
49	G	301.19	151.10	284.16	142.58			3
02	P	244.17	122.59	227.14	114.07			2
	K	147.11	74.06	130.09	65.55			1

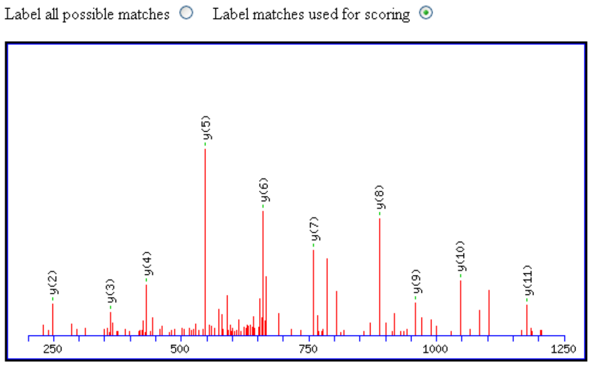
MASCOT : *Search Results FAQ*

MATRIX SCIENCE

In 2005, we explained that the peaks labelled in a Peptide View report are those used for scoring. Mascot begins by selecting a small number of experimental peaks on the basis of normalised intensity. It calculates a probability based score according to the number of matches. It then increases the number of selected peaks and recalculates the score. It continues to iterate until it is clear that the score can only get worse. It then reports the best score it found, which should correspond to an optimum selection, taking mostly real peaks and leaving behind mostly noise.

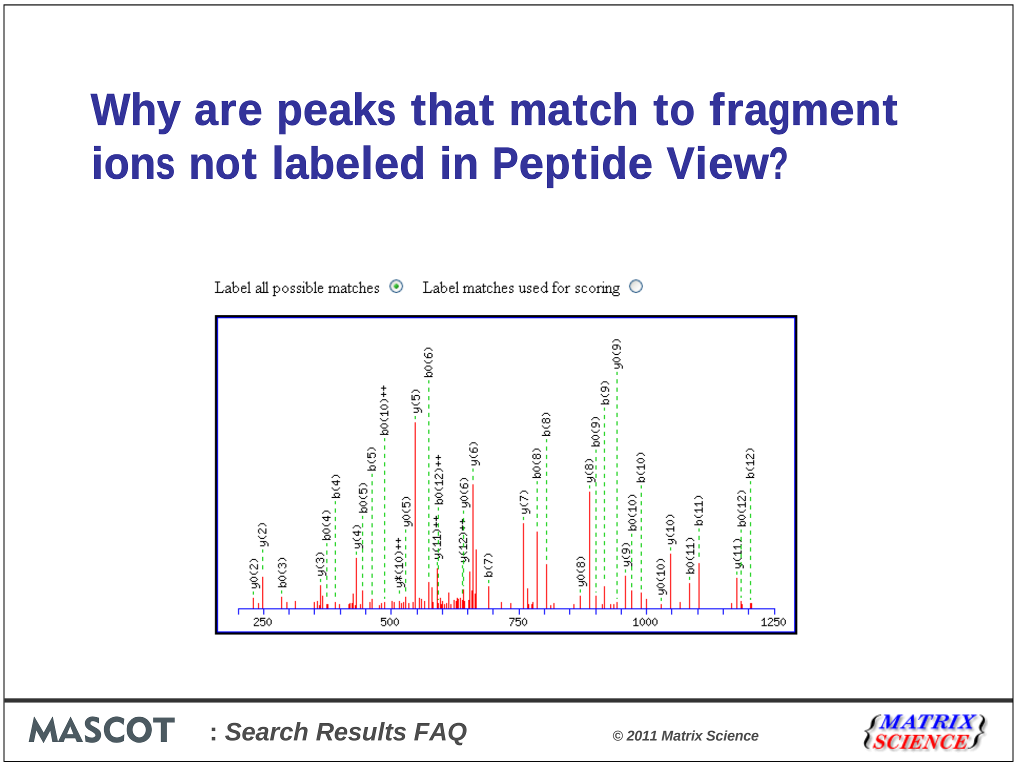
Mascot is not trying to find all possible matches in the spectrum. As in this example, many spectra have "peak at every mass" noise, and can match any fragment ion from any sequence if there is no intensity discrimination. So, you may look at a peptide view report and see obvious matches that are unlabelled. However, if the peak selection was to be extended to include these additional matches, it would also have to include a number of additional noise peaks, and the score would decrease.

Why are peaks that match to fragment ions not labeled in Peptide View?



Label all possible matches ○ Label matches used for scoring ●

You finally wore us down. After 10 years of giving this reply, we gave up arguing and made it a user option. The default is to label the peaks used for scoring



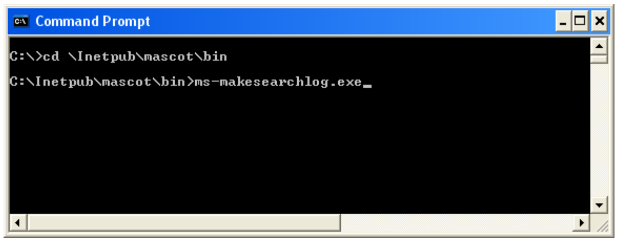
Why are peaks that match to fragment ions not labeled in Peptide View?

Label all possible matches ● Label matches used for scoring ○

MASCOT : Search Results FAQ © 2011 Matrix Science MATRIX SCIENCE

And the radio buttons make it easy to switch to having all possible matches labelled.

I’ve accidently deleted the search log, what can I do?



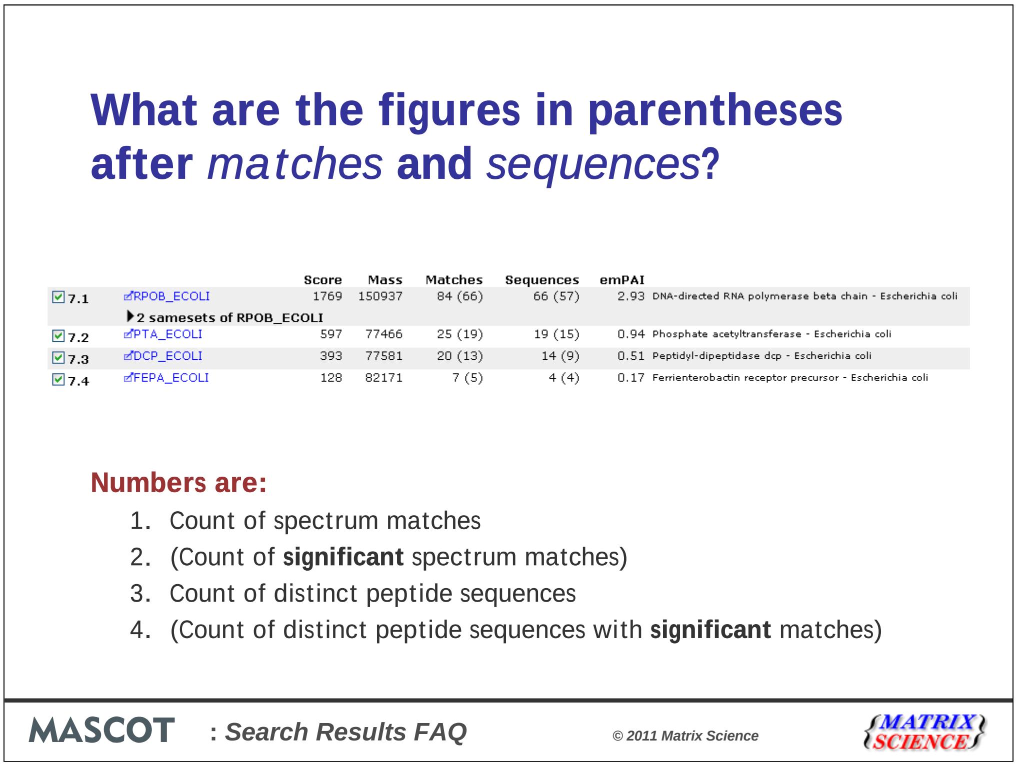
```
C:\>cd \Inetpub\mascot\bin

C:\Inetpub\mascot\bin>ms-makesearchlog.exe_
```

Also useful for pruning the search log after deleting unwanted result files

Now for the new questions!

Losing the search log isn’t a disaster. Mascot includes a utility that crawls the files under the data directory and re-creates the log file, searches.log. The utility is called ms-makesearchlog.exe and lives in the Mascot bin directory. If there are a lot of result files, this can take a long time, so maybe run it over the weekend. This can also be useful if you’ve deleted unwanted result files and want to create a clean search log



What are the figures in parentheses after *matches* and *sequences*?

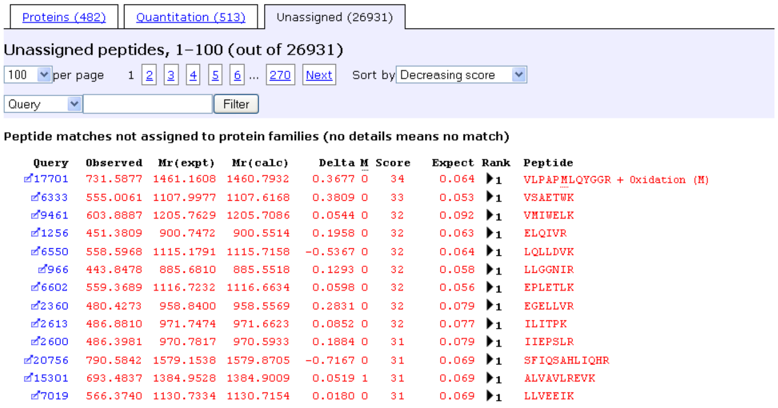
		Score	Mass	Matches	Sequences	emPAI	
☑ 7.1	RPOB_ECOLI	1769	150937	84 (66)	66 (57)	2.93	DNA-directed RNA polymerase beta chain - Escherichia coli
	▶ 2 samesets of RPOB_ECOLI						
☑ 7.2	PTA_ECOLI	597	77466	25 (19)	19 (15)	0.94	Phosphate acetyltransferase - Escherichia coli
☑ 7.3	DCP_ECOLI	393	77581	20 (13)	14 (9)	0.51	Peptidyl-dipeptidase dcp - Escherichia coli
☑ 7.4	FEPA_ECOLI	128	82171	7 (5)	4 (4)	0.17	Ferrienterobactin receptor precursor - Escherichia coli

Numbers are:

1. Count of spectrum matches
2. (Count of **significant** spectrum matches)
3. Count of distinct peptide sequences
4. (Count of distinct peptide sequences with **significant** matches)

The first number under matches is the count of MS/MS spectra that have been matched to peptides from this protein. The number in brackets is the count of significant matches. That is, matches with scores above the significance threshold. The number under sequences is the count of matches to distinct peptide sequences. If multiple spectra match to the same sequence, even if the charge state or modification state is different, this only counts as one sequence. Again, the number in brackets is the count of sequences with significant matches. When making a list of protein IDs, its often advisable to have significant matches to at least two distinct sequences before accepting a protein ID as safe.

Where do the peptide sequences come from in the unassigned list?



Unassigned peptides, 1–100 (out of 26931)

Peptide matches not assigned to protein families (no details means no match)

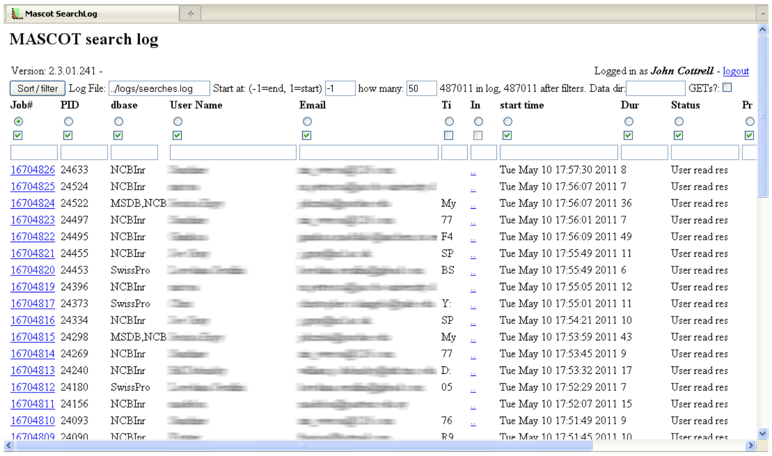
Query	Observed	Mr(expt)	Mr(calc)	Delta	M	Score	Expect	Rank	Peptide
17701	731.5877	1461.1608	1460.7932	0.3677	0	34	0.064	▶1	VLPAPMLQYGGR + Oxidation (M)
6333	555.0061	1107.9977	1107.6168	0.3809	0	33	0.053	▶1	VSAETWK
9461	603.8887	1205.7629	1205.7086	0.0544	0	32	0.092	▶1	VMIWELK
1256	451.3809	900.7472	900.5514	0.1958	0	32	0.063	▶1	ELQIVR
6550	558.5968	1115.1791	1115.7158	-0.5367	0	32	0.064	▶1	LQLLDVK
966	443.8478	885.6810	885.5518	0.1293	0	32	0.058	▶1	LLGGNIR
6602	559.3689	1116.7232	1116.6634	0.0598	0	32	0.056	▶1	EPLETLK
2360	480.4273	958.8400	958.5569	0.2831	0	32	0.079	▶1	EGELLVR
2613	486.8810	971.7474	971.6623	0.0852	0	32	0.077	▶1	ILITPK
2600	486.3981	970.7817	970.5933	0.1884	0	31	0.079	▶1	IIEPSLR
20756	790.5842	1579.1538	1579.8705	-0.7167	0	31	0.069	▶1	SFIQSAHLIQHR
15301	693.4837	1384.9528	1384.9009	0.0519	1	31	0.069	▶1	ALVAVLREVK
7019	566.3740	1130.7334	1130.7154	0.0180	0	31	0.069	▶1	LLVEEIK

The way this question is worded indicates a common misunderstanding. Unassigned doesn't mean unmatched. The unassigned list contains peptide matches that are not assigned to proteins in the body of the report. Its the overflow, if you like. If you reformat the report, asking for more and more protein hits, all of the unassigned matches will eventually get pulled into the body of the report. Conversely, if you specify that you only want a single protein hit for a result which should have many hits, you will find a lot of high scoring, significant matches pushed onto the unassigned list.

There may be some unmatched spectra on the unassigned list, but these are the entries with nothing apart from the experimental m/z and Mr values.

When you load a Peptide View report by clicking on a query number in the unassigned list, it refers to the first protein containing the matched peptide. This may or may not be the protein that would be the anchor protein if more hits were specified and the peptide match was pulled into the body of the report.

How can I get hold of the raw result file?



MASCOT search log

Version: 2.3.01.241 -

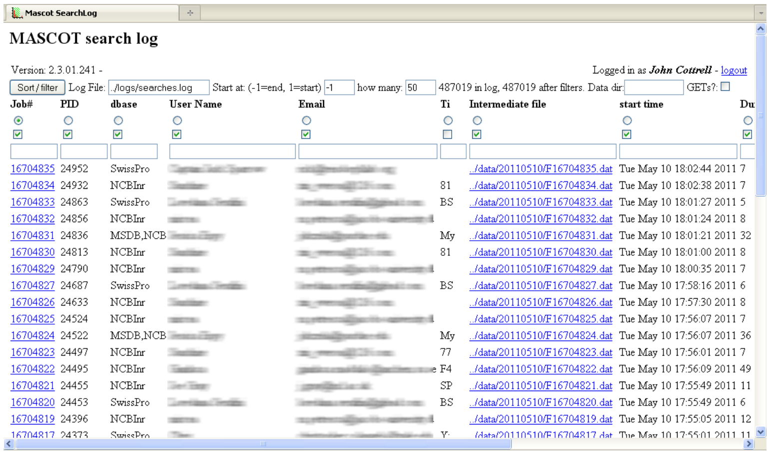
Logged in as ***John Cottrell*** - logout

Sort / filter Log File: ../logs/searches.log Start at: (-1=end, 1=start) -1 how many: 50 487011 in log, 487011 after filters. Data dir: GETs?: ☐

Job#	PID	dbase	User Name	Email	Ti	In	start time	Dur	Status	Pr
●	○	○	○	○	○	○	○	○	○	○
☑	☑	☑	☑	☑	☐	☐	☑	☑	☑	☑
16704826	24633	NCBInr	[redacted]	[redacted]		..	Tue May 10 17:57:30 2011	8	User read res	
16704825	24524	NCBInr	[redacted]	[redacted]		..	Tue May 10 17:56:07 2011	7	User read res	
16704824	24522	MSDB,NCB	[redacted]	[redacted]	My	..	Tue May 10 17:56:07 2011	36	User read res	
16704823	24497	NCBInr	[redacted]	[redacted]	77	..	Tue May 10 17:56:01 2011	7	User read res	
16704822	24495	NCBInr	[redacted]	[redacted]	F4	..	Tue May 10 17:56:09 2011	49	User read res	
16704821	24455	NCBInr	[redacted]	[redacted]	SP	..	Tue May 10 17:55:49 2011	11	User read res	
16704820	24453	SwissPro	[redacted]	[redacted]	BS	..	Tue May 10 17:55:49 2011	6	User read res	
16704819	24396	NCBInr	[redacted]	[redacted]		..	Tue May 10 17:55:05 2011	12	User read res	
16704817	24373	SwissPro	[redacted]	[redacted]	Y:	..	Tue May 10 17:55:01 2011	11	User read res	
16704816	24334	NCBInr	[redacted]	[redacted]	SP	..	Tue May 10 17:54:21 2011	10	User read res	
16704815	24298	MSDB,NCB	[redacted]	[redacted]	My	..	Tue May 10 17:53:59 2011	43	User read res	
16704814	24269	NCBInr	[redacted]	[redacted]	77	..	Tue May 10 17:53:45 2011	9	User read res	
16704813	24240	NCBInr	[redacted]	[redacted]	D:	..	Tue May 10 17:53:32 2011	17	User read res	
16704812	24180	SwissPro	[redacted]	[redacted]	05	..	Tue May 10 17:52:29 2011	7	User read res	
16704811	24156	NCBInr	[redacted]	[redacted]		..	Tue May 10 17:52:07 2011	15	User read res	
16704810	24093	NCBInr	[redacted]	[redacted]	76	..	Tue May 10 17:51:49 2011	9	User read res	
16704809	24090	NCBInr	[redacted]	[redacted]	R9		Tue May 10 17:51:45 2011	10	User read res	

Some software packages, such as Scaffold, work off the Mascot result file. If you don’t have file share access to the Mascot server, there are two ways to download a result file. One is from the link in the search log. If you’ve ever wondered what the mysterious column headed ‘In’ containing two dots was, try checking the box at the top and reloading.

How can I get hold of the raw result file?



MASCOT search log

Version: 2.3.01.241 -

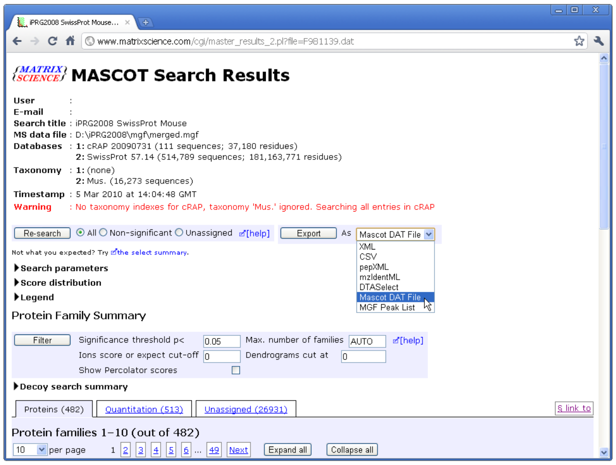
Logged in as ***John Cottrell*** - logout

Log File: ../logs/searches.log Start at: (-1=end, 1=start) -1 how many: 50 487019 in log, 487019 after filters. Data dir: GETs?: ☐

Job#	PID	dbase	User Name	Email	Ti	Intermediate file	start time	Du
●	○	○	○	○	○	○	○	○
☑	☑	☑	☑	☑	☐	☑	☑	☑
16704835	24952	SwissPro	[redacted]	[redacted]		../data/20110510/F16704835.dat	Tue May 10 18:02:44 2011	7
16704834	24932	NCBInr	[redacted]	[redacted]	81	../data/20110510/F16704834.dat	Tue May 10 18:02:38 2011	7
16704833	24863	SwissPro	[redacted]	[redacted]	BS	../data/20110510/F16704833.dat	Tue May 10 18:01:27 2011	5
16704832	24856	NCBInr	[redacted]	[redacted]		../data/20110510/F16704832.dat	Tue May 10 18:01:24 2011	8
16704831	24836	MSDB,NCB	[redacted]	[redacted]	My	../data/20110510/F16704831.dat	Tue May 10 18:01:21 2011	32
16704830	24813	NCBInr	[redacted]	[redacted]	81	../data/20110510/F16704830.dat	Tue May 10 18:01:00 2011	8
16704829	24790	NCBInr	[redacted]	[redacted]		../data/20110510/F16704829.dat	Tue May 10 18:00:35 2011	7
16704827	24687	SwissPro	[redacted]	[redacted]	BS	../data/20110510/F16704827.dat	Tue May 10 17:58:16 2011	6
16704826	24633	NCBInr	[redacted]	[redacted]		../data/20110510/F16704826.dat	Tue May 10 17:57:30 2011	8
16704825	24524	NCBInr	[redacted]	[redacted]		../data/20110510/F16704825.dat	Tue May 10 17:56:07 2011	7
16704824	24522	MSDB,NCB	[redacted]	[redacted]	My	../data/20110510/F16704824.dat	Tue May 10 17:56:07 2011	36
16704823	24497	NCBInr	[redacted]	[redacted]	77	../data/20110510/F16704823.dat	Tue May 10 17:56:01 2011	7
16704822	24495	NCBInr	[redacted]	[redacted]	F4	../data/20110510/F16704822.dat	Tue May 10 17:56:09 2011	49
16704821	24455	NCBInr	[redacted]	[redacted]	SP	../data/20110510/F16704821.dat	Tue May 10 17:55:49 2011	11
16704820	24453	SwissPro	[redacted]	[redacted]	BS	../data/20110510/F16704820.dat	Tue May 10 17:55:49 2011	6
16704819	24396	NCBInr	[redacted]	[redacted]		../data/20110510/F16704819.dat	Tue May 10 17:55:05 2011	12
16704817	24373	SwissPro	[redacted]	[redacted]	Y:	../data/20110510/F16704817.dat	Tue May 10 17:55:01 2011	11

Some columns are displayed in full by default, others are collapsed to the width of a checkbox, and this is one of them. The full header is 'Intermediate file' (don't ask me why) and the contents are links to the result files

How can I get hold of the raw result file?



MASCOT Search Results

User :
E-mail :
Search title : iPRG2008 SwissProt Mouse
MS data file : D:\iPRG2008\mgf\merged.mgf
Databases : **1:** cRAP 20090731 (111 sequences; 37,180 residues)
2: SwissProt 57.14 (514,789 sequences; 181,163,771 residues)
Taxonomy : **1:** (none)
2: Mus. (16,273 sequences)
Timestamp : 5 Mar 2010 at 14:04:48 GMT
Warning : No taxonomy indexes for cRAP, taxonomy 'Mus.' ignored. Searching all entries in cRAP

Not what you expected? Try the select summary.

Search parameters
Score distribution
Legend

Protein Family Summary

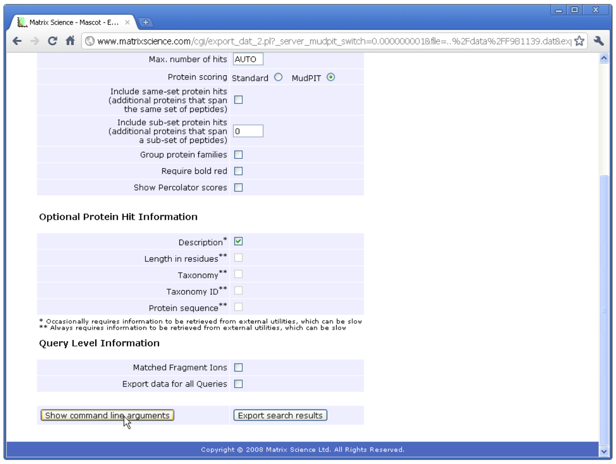
Significance threshold p< 0.05
Max. number of families AUTO
Ions score or expect cut-off 0
Dendrograms cut at 0
Show Percolator scores ☐

Decoy search summary

Protein families 1–10 (out of 482)

If you don't have access to the search log, and you are running Mascot 2.3, you can also get a result file by loading the result report in a browser and choosing Mascot DAT file as the export format

How can I export a very large search result without a time-out?



Max. number of hits AUTO

Protein scoring Standard ○ MudPIT ●

Include same-set protein hits (additional proteins that span the same set of peptides) ☐

Include sub-set protein hits (additional proteins that span a sub-set of peptides) 0

Group protein families ☐

Require bold red ☐

Show Percolator scores ☐

Optional Protein Hit Information

Description* ☑

Length in residues** ☐

Taxonomy** ☐

Taxonomy ID** ☐

Protein sequence** ☐

* Occasionally requires information to be retrieved from external utilities, which can be slow
** Always requires information to be retrieved from external utilities, which can be slow

Query Level Information

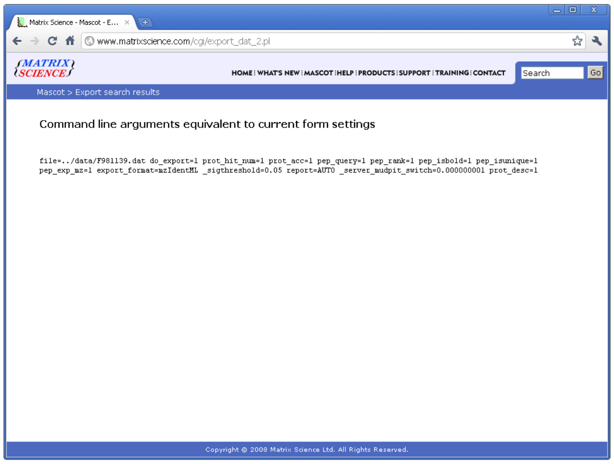
Matched Fragment Ions ☐

Export data for all Queries ☐

The topic of exporting brings us to our next question: How can I export a very large search result without a time-out?

Processing a large result to create some of the export formats takes so long that the browser connection times out. We should make the script more sophisticated, and have it do something to keep the connection alive until the file is ready to download, but right now the fix is to perform the export at the command line. To make this easy, you can ask the export script to display the arguments corresponding to the options in the export form

How can I export a very large search result without a time-out?

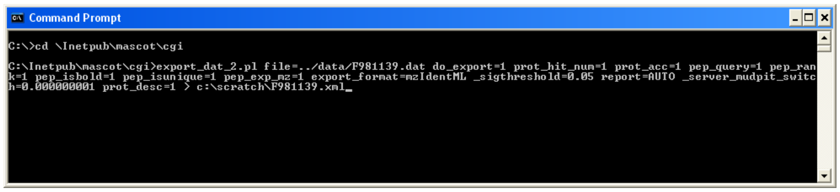


Command line arguments equivalent to current form settings

```
file=../data/F981139.dat do_export=1 prot_hit_num=1 prot_acc=1 pep_query=1 pep_rank=1 pep_isbold=1 pep_isunique=1
pep_exp_mz=1 export_format=mzIdentML _sigthreshold=0.05 report=AUTO _server_mudpit_switch=0.000000001 prot_desc=1
```

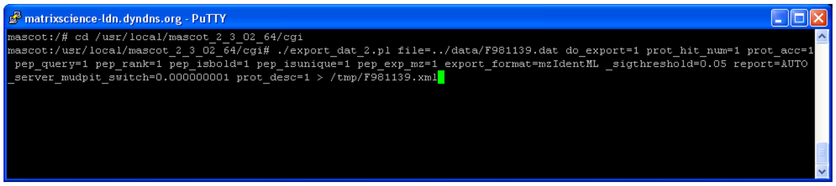
Here’s an example

How can I export a very large search result without a time-out?



```
C:\>cd \Inetpub\mascot\cgi

C:\Inetpub\mascot\cgi>export_dat_2.pl file=../data/F981139.dat do_export=1 prot_hit_num=1 prot_acc=1 pep_query=1 pep_ran
k=1 pep_isbold=1 pep_isunique=1 pep_exp_mz=1 export_format=mzIdentML _sigthreshold=0.05 report=AUTO _server_mudpit_switc
h=0.000000001 prot_desc=1 > c:\scratch\F981139.xml_
```

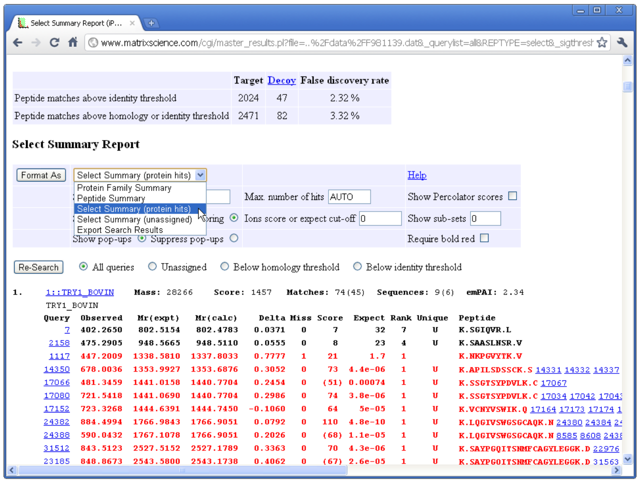


```
mascot:/# cd /usr/local/mascot_2_3_02_64/cgi
mascot:/usr/local/mascot_2_3_02_64/cgi# ./export_dat_2.pl file=../data/F981139.dat do_export=1 prot_hit_num=1 prot_acc=1
 pep_query=1 pep_rank=1 pep_isbold=1 pep_isunique=1 pep_exp_mz=1 export_format=mzIdentML _sigthreshold=0.05 report=AUTO
_server_mudpit_switch=0.000000001 prot_desc=1 > /tmp/F981139.xml
```

Open a command window and change to the Mascot cgi directory. Copy and paste the arguments after the name of the script and pipe the output to a file name.

The command line is also the way to go if you need to export a large number of files. Use a text editor or Excel to create a file with a line such as this for each of the result files. Save and execute as a DOS batch file in Windows or a shell script in Linux. It is also possible to write a short script that takes the result file path as an argument and executes the export script, as here. This can be called from Mascot Daemon as an external process to export search results automatically. Just email support@matrixscience.com if you want more details.

How can I get the old-style result report that my software requires?



	Target	Decoy	False discovery rate
Peptide matches above identity threshold	2024	47	2.32 %
Peptide matches above homology or identity threshold	2471	82	3.32 %

Select Summary Report

Format As: Select Summary (protein hits)

Max. number of hits AUTO

Show Percolator scores ☐

Ions score or expect cut-off 0

Show sub-sets 0

Show pop-ups ● Suppress pop-ups ○

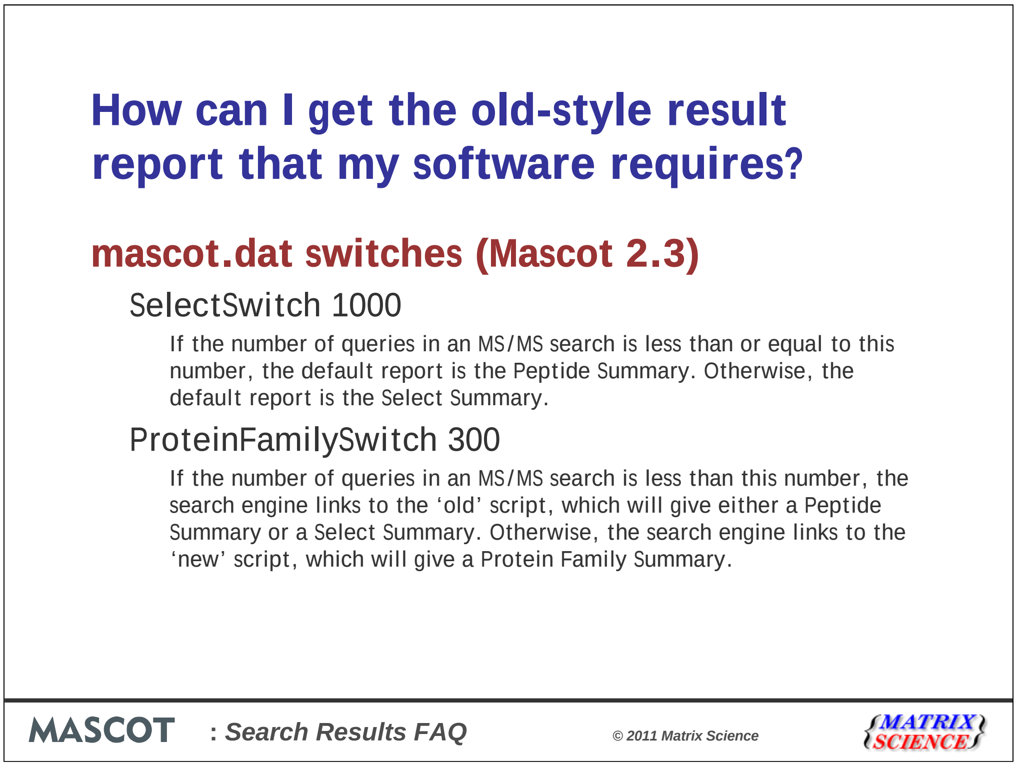
Require bold red ☐

● All queries ○ Unassigned ○ Below homology threshold ○ Below identity threshold

```
1.     1::TRY1_BOVIN    Mass: 28266    Score: 1457    Matches: 74(45)   Sequences: 9(6)   emPAI: 2.34
       TRY1_BOVIN
      Query  Observed   Mr(expt)   Mr(calc)    Delta Miss Score  Expect Rank Unique  Peptide
          7  402.2650   802.5154   802.4783   0.0371    0     7      32    7      U  K.SGIQVR.L
       2158  475.2905   948.5665   948.5110   0.0555    0     8      23    4      U  K.SAASLNSR.V
       1117  447.2009  1338.5810  1337.8033   0.7777    1    21     1.7    1         K.NKPGVYTK.V
      14350  678.0036  1353.9927  1353.6876   0.3052    0    73 4.4e-06    1      U  K.APILSDSSCK.S 14331 14332 14337
      17066  481.3459  1441.0158  1440.7704   0.2454    0  (51) 0.00074    1      U  K.SSGTSYPDVLK.C 17067
      17080  721.5418  1441.0690  1440.7704   0.2986    0    74 3.8e-06    1      U  K.SSGTSYPDVLK.C 17034 17042
      17152  723.3268  1444.6391  1444.7450  -0.1060    0    64   5e-05    1      U  K.VCNYVSWIK.Q 17164 17173 17174
      24382  884.4994  1766.9843  1766.9051   0.0792    0   110 4.8e-10    1      U  K.LQGIVSWGSGCAQK.N 24380 24384
      24388  590.0432  1767.1078  1766.9051   0.2026    0  (68) 1.1e-05    1      U  K.LQGIVSWGSGCAQK.N 8585 8608
      31512  843.5123  2527.5152  2527.1789   0.3363    0    70 4.3e-06    1      U  K.SAYPGQITSNMFCAGYLEGGK.D 22976
      23185  848.8673  2543.5800  2543.1738   0.4062    0  (67) 2.6e-05    1      U  K.SAYPGQITSNMFCAGYLEGGK.D 31563
```

If someone is developing a software application that processes Mascot results, we always advise using Mascot Parser to take the information from the result file. However, there are applications out there that try to ‘screen scrape’ the search results from the HTML report. When the report format changes, such applications are likely to break. So, we sometimes get asked whether it is possible to create a Mascot 2.2 - style report in Mascot 2.3.

In most cases, the problem is that software assumes that the report will be a particular format, such as a Peptide Summary, even though Mascot switches between formats according to the type of search. If the application requires you to save the report to a file from the web browser, you can easily fix this by selecting the required format in the controls near the top of the report.



How can I get the old-style result report that my software requires?

mascot.dat switches (Mascot 2.3)

SelectSwitch 1000

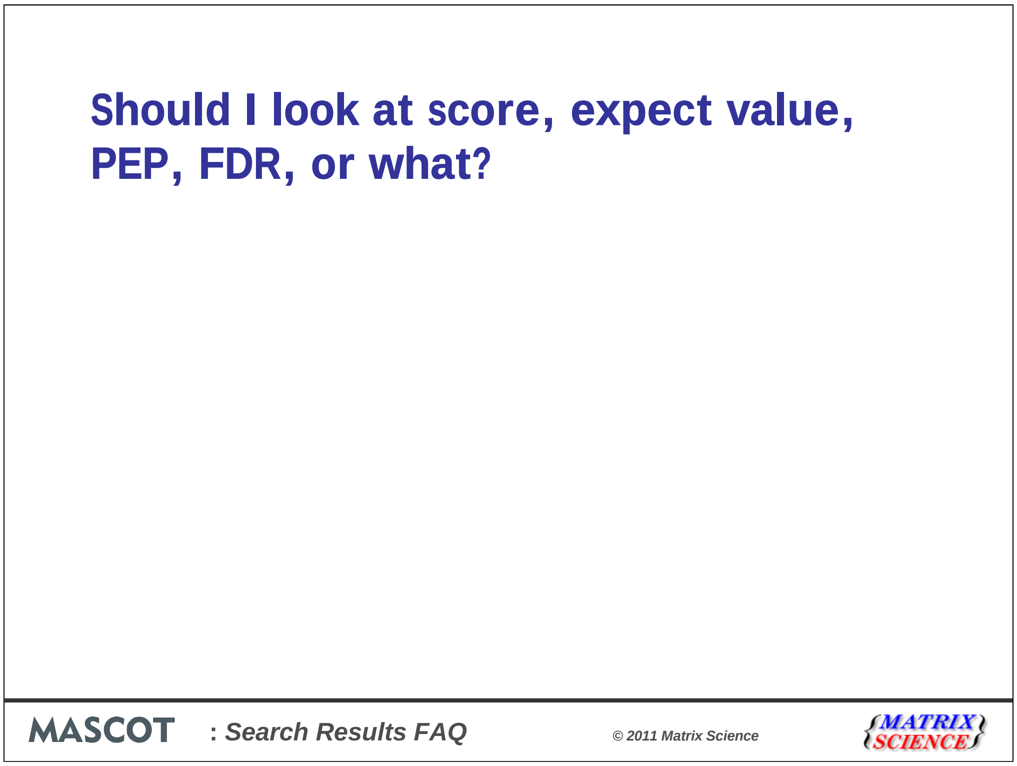
If the number of queries in an MS/MS search is less than or equal to this number, the default report is the Peptide Summary. Otherwise, the default report is the Select Summary.

ProteinFamilySwitch 300

If the number of queries in an MS/MS search is less than this number, the search engine links to the ‘old’ script, which will give either a Peptide Summary or a Select Summary. Otherwise, the search engine links to the ‘new’ script, which will give a Protein Family Summary.

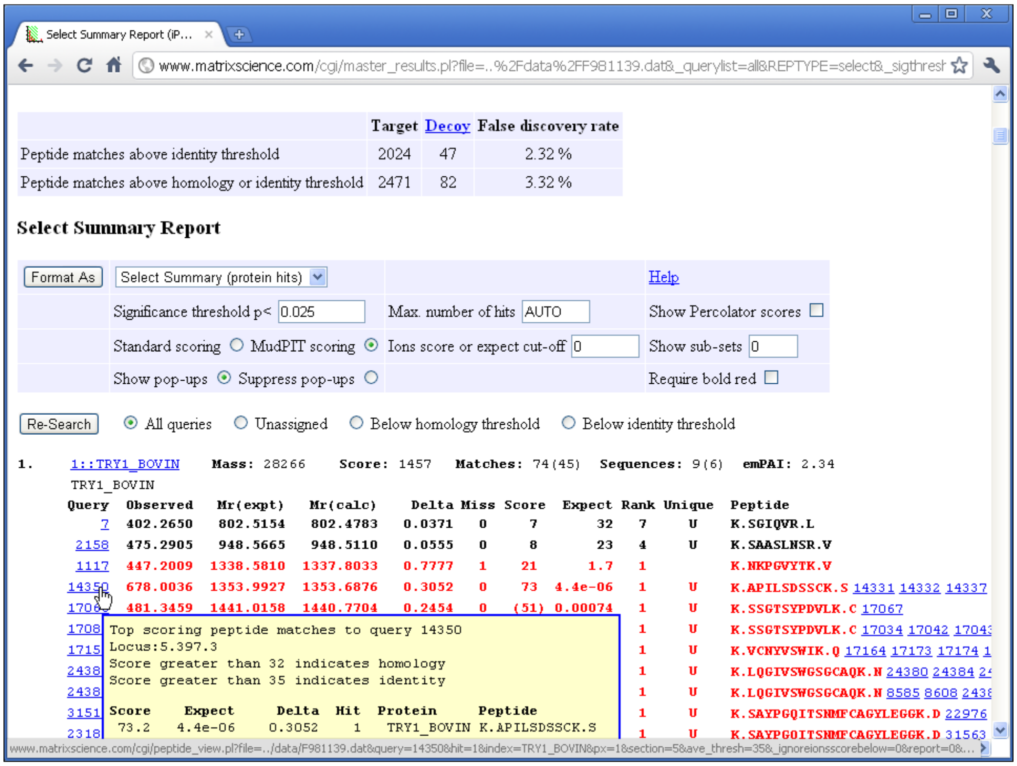
MASCOT : Search Results FAQ

If the software tries to capture the result report HTML directly, you may not have this option. The other approach is to change the settings in mascot.dat so that the required report type becomes the default in all cases. These switches vary between versions, as new report types are added. There is quite a lot of complexity to this, and you need to read the manual to get a complete picture, but if you wanted to get the Peptide Summary for all MS/MS searches, you would set ProteinFamilySwitch to a high value, e.g. 1 million, to ensure the ‘old’ script (master_results.pl) was always called at the end of a search. You would also set SelectSwitch to a high value to ensure that the report format defaulted to Peptide Summary. In general, this is not a good idea, because you are likely to run into time-out or memory problems with large searches. But, it may be the only way to get the results into an old software application.



Should I look at score, expect value, PEP, FDR, or what?

Or, in other words, what do all these numbers mean?



	Target	Decoy	False discovery rate
Peptide matches above identity threshold	2024	47	2.32 %
Peptide matches above homology or identity threshold	2471	82	3.32 %

Select Summary Report

Select Summary (protein hits)

Significance threshold p< 0.025

Max. number of hits AUTO

Show Percolator scores ☐

Standard scoring ○ MudPIT scoring ●

Ions score or expect cut-off 0

Show sub-sets 0

Show pop-ups ● Suppress pop-ups ○

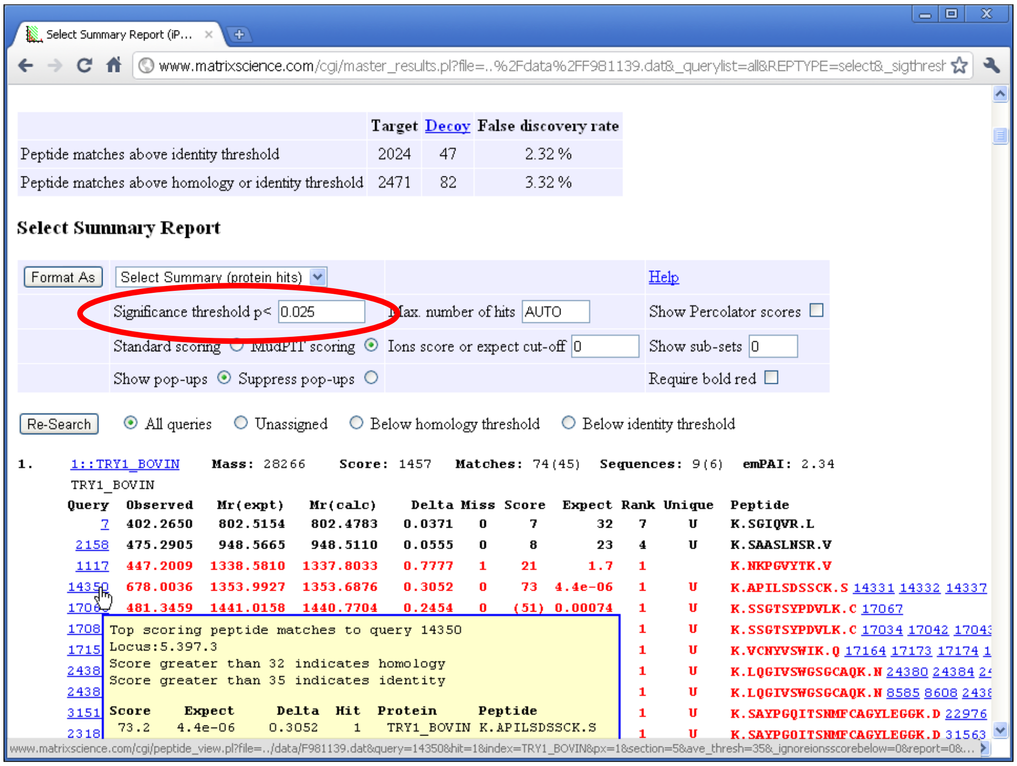
Require bold red ☐

● All queries ○ Unassigned ○ Below homology threshold ○ Below identity threshold

```
1.     1::TRY1_BOVIN     Mass: 28266     Score: 1457    Matches: 74(45)   Sequences: 9(6)   emPAI: 2.34
       TRY1_BOVIN
      Query  Observed   Mr(expt)   Mr(calc)    Delta Miss Score  Expect Rank Unique  Peptide
          7  402.2650   802.5154   802.4783   0.0371   0     7      32    7     U    K.SGIQVR.L
       2158  475.2905   948.5665   948.5110   0.0555   0     8      23    4     U    K.SAASLNSR.V
       1117  447.2009  1338.5810  1337.8033   0.7777   1    21     1.7    1          K.NKPGVYTK.V
      14350  678.0036  1353.9927  1353.6876   0.3052   0    73  4.4e-06   1     U    K.APILSDSSCK.S 14331 14332 14337
      170    481.3459  1441.0158  1440.7704   0.2454   0  (51) 0.00074   1     U    K.SSGTSYPDVLK.C 17067
      1708 Top scoring peptide matches to query 14350                      1     U    K.SSGTSYPDVLK.C 17034 17042 1704
      1715 Locus:5.397.3                                                   1     U    K.VCNYVSWIK.Q 17164 17173 17174 1
      2438 Score greater than 32 indicates homology                        1     U    K.LQGIVSWGSGCAQK.N 24380 24384 24
      2438 Score greater than 35 indicates identity                        1     U    K.LQGIVSWGSGCAQK.N 8585 8608 2438
      3151 Score      Expect      Delta   Hit   Protein      Peptide       1     U    K.SAYPGQITSNMFCAGYLEGGK.D 22976
      2318  73.2     4.4e-06     0.3052    1   TRY1_BOVIN K.APILSDSSCK.S   1     U    K.SAYPGQITSNMFCAGYLEGGK.D 31563
```

www.matrixscience.com/cgi/peptide_view.pl?file=../data/F981139.dat&query=14350&hit=1&index=TRY1_BOVIN&px=1&section=5&ave_thresh=35&_ignoreionsscorebelow=0&report=0&…

I’ll be the first to admit that there are a confusing number of ways to measure the quality of search results.



	Target	Decoy	False discovery rate
Peptide matches above identity threshold	2024	47	2.32 %
Peptide matches above homology or identity threshold	2471	82	3.32 %

Select Summary Report

Significance threshold p< 0.025
Max. number of hits AUTO
Show Percolator scores ☐
Standard scoring ○ MudPIT scoring ●
Ions score or expect cut-off 0
Show sub-sets 0
Show pop-ups ● Suppress pop-ups ○
Require bold red ☐

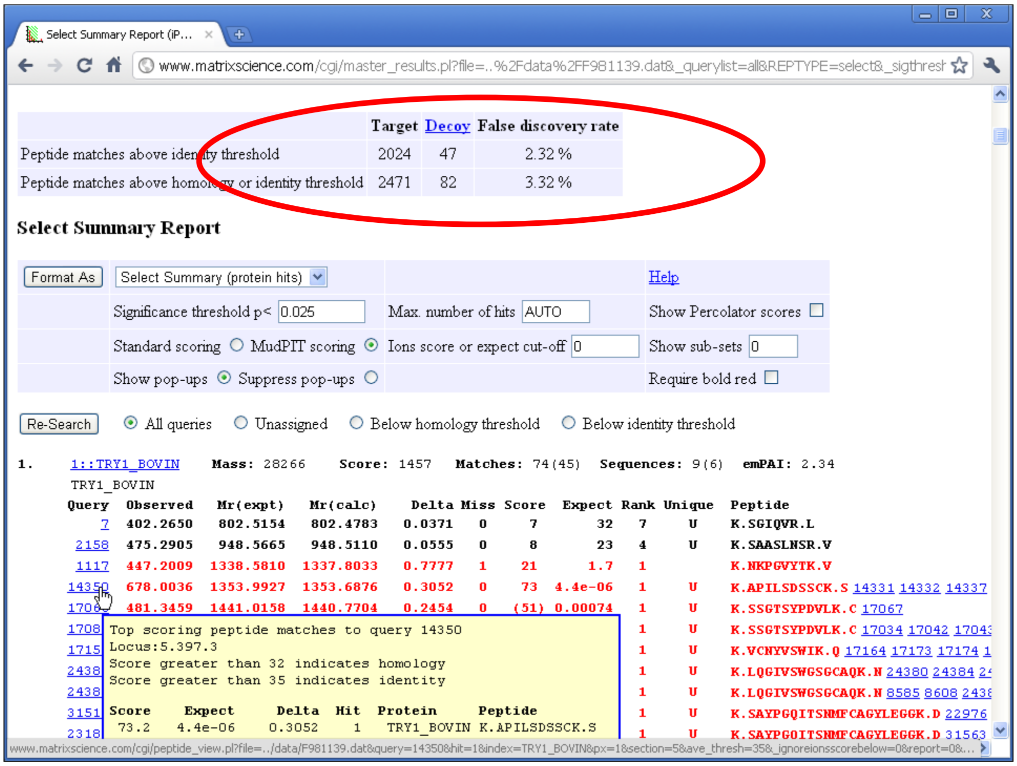
● All queries ○ Unassigned ○ Below homology threshold ○ Below identity threshold

```
1.     1::TRY1_BOVIN     Mass: 28266     Score: 1457    Matches: 74(45)   Sequences: 9(6)   emPAI: 2.34
       TRY1_BOVIN
      Query  Observed   Mr(expt)   Mr(calc)    Delta Miss Score  Expect Rank Unique  Peptide
          7  402.2650   802.5154   802.4783   0.0371    0     7      32    7     U    K.SGIQVR.L
       2158  475.2905   948.5665   948.5110   0.0555    0     8      23    4     U    K.SAASLNSR.V
       1117  447.2009  1338.5810  1337.8033   0.7777    1    21     1.7    1          K.NKPGVYTK.V
      14350  678.0036  1353.9927  1353.6876   0.3052    0    73  4.4e-06   1     U    K.APILSDSSCK.S 14331 14332 14337
       170   481.3459  1441.0158  1440.7704   0.2454    0  (51) 0.00074   1     U    K.SSGTSYPDVLK.C 17067
       1708                                                                 1     U    K.SSGTSYPDVLK.C 17034 17042 1704
       1715                                                                 1     U    K.VCNYVSWIK.Q 17164 17173 17174 1
       2438                                                                 1     U    K.LQGIVSWGSGCAQK.N 24380 24384 24
       2438                                                                 1     U    K.LQGIVSWGSGCAQK.N 8585 8608 243
       3151                                                                 1     U    K.SAYPGQITSNMFCAGYLEGGK.D 22976
       2318                                                                 1     U    K.SAYPGQITSNMFCAGYLEGGK.D 31563
```

```
Top scoring peptide matches to query 14350
Locus:5.397.3
Score greater than 32 indicates homology
Score greater than 35 indicates identity

Score     Expect     Delta   Hit   Protein     Peptide
 73.2    4.4e-06    0.3052    1    TRY1_BOVIN K.APILSDSSCK.S
```

The significance threshold is a goal. It tells Mascot the approximate false discovery rate we would like to see. In this case, 2.5%. To find the actual false discovery rate, we need to run a target decoy search. Hopefully, we'll find that the numbers are reasonably close to the goal. If not, we can adjust the significance threshold to achieve the desired false discovery rate by trial and error.



	Target	Decoy	False discovery rate
Peptide matches above identity threshold	2024	47	2.32 %
Peptide matches above homology or identity threshold	2471	82	3.32 %

Select Summary Report

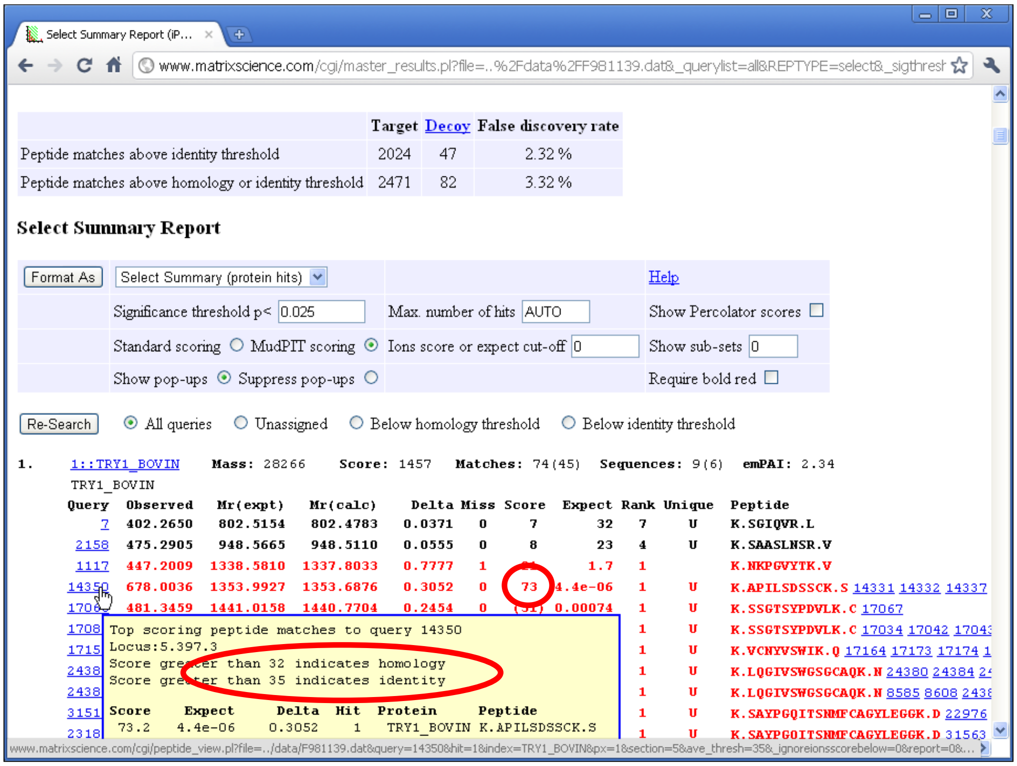
Format As	Select Summary (protein hits)		Help
	Significance threshold p< 0.025	Max. number of hits AUTO	Show Percolator scores ☐
	Standard scoring ○ MudPIT scoring ●	Ions score or expect cut-off 0	Show sub-sets 0
	Show pop-ups ● Suppress pop-ups ○		Require bold red ☐

Re-Search ● All queries ○ Unassigned ○ Below homology threshold ○ Below identity threshold

```
1.     1::TRY1_BOVIN     Mass: 28266     Score: 1457    Matches: 74(45)   Sequences: 9(6)   emPAI: 2.34
       TRY1_BOVIN
      Query  Observed    Mr(expt)    Mr(calc)     Delta Miss Score   Expect Rank Unique   Peptide
          7  402.2650    802.5154    802.4783    0.0371    0     7       32    7      U    K.SGIQVR.L
       2158  475.2905    948.5665    948.5110    0.0555    0     8       23    4      U    K.SAASLNSR.V
       1117  447.2009   1338.5810   1337.8033    0.7777    1    21      1.7    1           K.NKPGVYTK.V
      14350  678.0036   1353.9927   1353.6876    0.3052    0    73  4.4e-06    1      U    K.APILSDSSCK.S 14331 14332 14337
      1706   481.3459   1441.0158   1440.7704    0.2454    0   (51) 0.00074   1      U    K.SSGTSYPDVLK.C 17067
      1708 Top scoring peptide matches to query 14350               1      U    K.SSGTSYPDVLK.C 17034 17042 1704
      1715 Locus:5.397.3                                             1      U    K.VCNYVSWIK.Q 17164 17173 17174 1
      2438 Score greater than 32 indicates homology                  1      U    K.LQGIVSWGSGCAQK.N 24380 24384 2
      2438 Score greater than 35 indicates identity                  1      U    K.LQGIVSWGSGCAQK.N 8585 8608 243
      3151 Score     Expect     Delta   Hit  Protein     Peptide      1      U    K.SAYPGQITSNMFCAGYLEGGK.D 22976
      2318  73.2     4.4e-06    0.3052    1   TRY1_BOVIN K.APILSDSSCK.S 1   U    K.SAYPGQITSNMFCAGYLEGGK.D 31563
```

Mascot calculates two score thresholds and reports a false discovery rate for each. The identity threshold is calculated from the number of trials. That is, the number of candidate peptide sequences in the database. Often, this is over-conservative, and the false discovery rate from the target decoy search will be lower than this. The homology threshold simply looks at whether the top match is an outlier from the distribution of random scores. Usually, this will give better sensitivity for a given false discovery rate. The main reason we still report the identity threshold is that you can't always calculate a homology threshold. Particularly when you combine very tight mass tolerances with a small database; there may be only one or two candidate peptides, so you can't tell whether the top match is an outlier. You can always calculate an identity threshold.

False discovery rate measures the quality of the search results as a whole. Here, it tells us that 2 or 3% of the matches with scores above threshold are likely to be wrong, but it doesn't tell us which ones.



	Target	Decoy	False discovery rate
Peptide matches above identity threshold	2024	47	2.32 %
Peptide matches above homology or identity threshold	2471	82	3.32 %

Select Summary Report

Format As: Select Summary (protein hits)
Help
Significance threshold p< 0.025
Max. number of hits AUTO
Show Percolator scores ☐
Standard scoring ○ MudPIT scoring ●
Ions score or expect cut-off 0
Show sub-sets 0
Show pop-ups ● Suppress pop-ups ○
Require bold red ☐

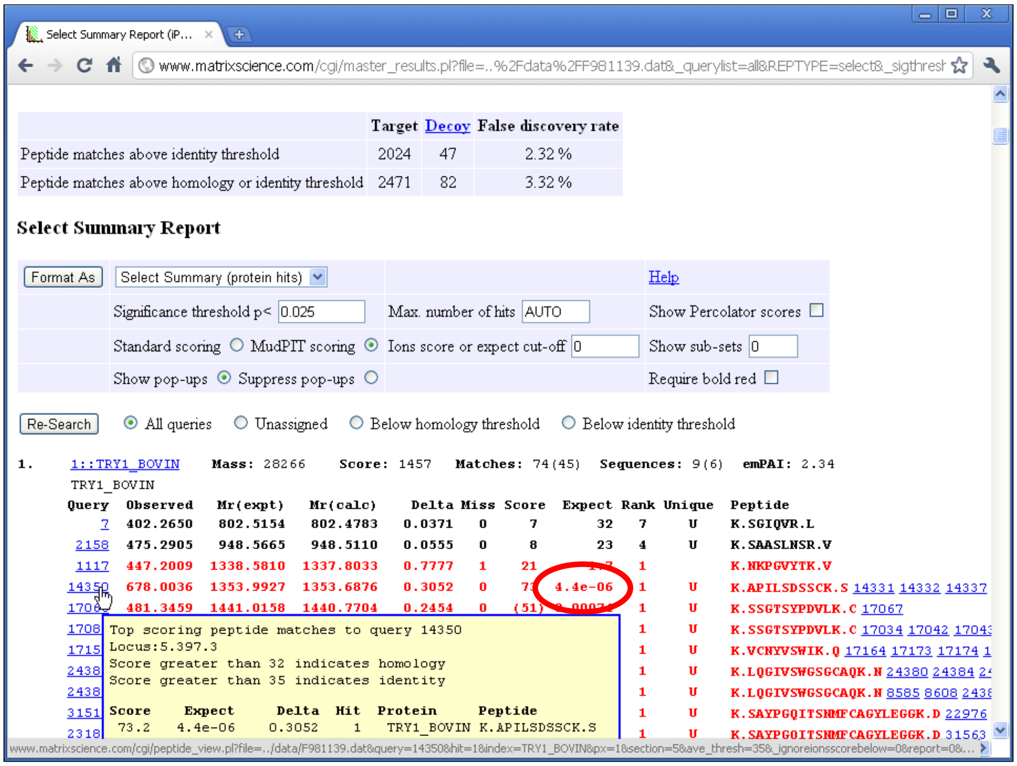
Re-Search ● All queries ○ Unassigned ○ Below homology threshold ○ Below identity threshold

```
1.      1::TRY1_BOVIN     Mass: 28266     Score: 1457    Matches: 74(45)   Sequences: 9(6)   emPAI: 2.34
        TRY1_BOVIN
       Query  Observed    Mr(expt)    Mr(calc)     Delta Miss Score  Expect Rank Unique   Peptide
           7  402.2650    802.5154    802.4783    0.0371    0     7      32    7     U    K.SGIQVR.L
        2158  475.2905    948.5665    948.5110    0.0555    0     8      23    4     U    K.SAASLNSR.V
        1117  447.2009   1338.5810   1337.8033    0.7777    1           1.7    1          K.NKPGVYTK.V
       14350  678.0036   1353.9927   1353.6876    0.3052    0    73 4.4e-06    1     U    K.APILSDSSCK.S 14331 14332 14337
        1706  481.3459   1441.0158   1440.7704    0.2454    0  (31) 0.00074    1     U    K.SSGTSYPDVLK.C 17067
        1708                                                                 1     U    K.SSGTSYPDVLK.C 17034 17042 1704
        1715                                                                 1     U    K.VCNYVSWIK.Q 17164 17173 17174 1
        2438                                                                 1     U    K.LQGIVSWGSGCAQK.N 24380 24384 2
        2438                                                                 1     U    K.LQGIVSWGSGCAQK.N 8585 8608 243
        3151                                                                 1     U    K.SAYPGQITSNMFCAGYLEGGK.D 22976
        2318                                                                 1     U    K.SAYPGQITSNMFCAGYLEGGK.D 31563
```

```
Top scoring peptide matches to query 14350
Locus:5.397.3
Score greater than 32 indicates homology
Score greater than 35 indicates identity

Score     Expect     Delta   Hit   Protein     Peptide
 73.2    4.4e-06    0.3052    1   TRY1_BOVIN K.APILSDSSCK.S
```

For each match, we have a Mascot score and one or two score thresholds, one for the identity threshold and one for the homology threshold. These scores are not fixed, they can be slightly different from match to match within a search and very different between searches. So, the score by itself doesn't tell you the quality of the match. You have to compare it with the chosen threshold score.



	Target	Decoy	False discovery rate
Peptide matches above identity threshold	2024	47	2.32 %
Peptide matches above homology or identity threshold	2471	82	3.32 %

Select Summary Report

Format As	Select Summary (protein hits)		Help
	Significance threshold p< 0.025	Max. number of hits AUTO	Show Percolator scores ☐
	Standard scoring ○ MudPIT scoring ●	Ions score or expect cut-off 0	Show sub-sets 0
	Show pop-ups ● Suppress pop-ups ○		Require bold red ☐

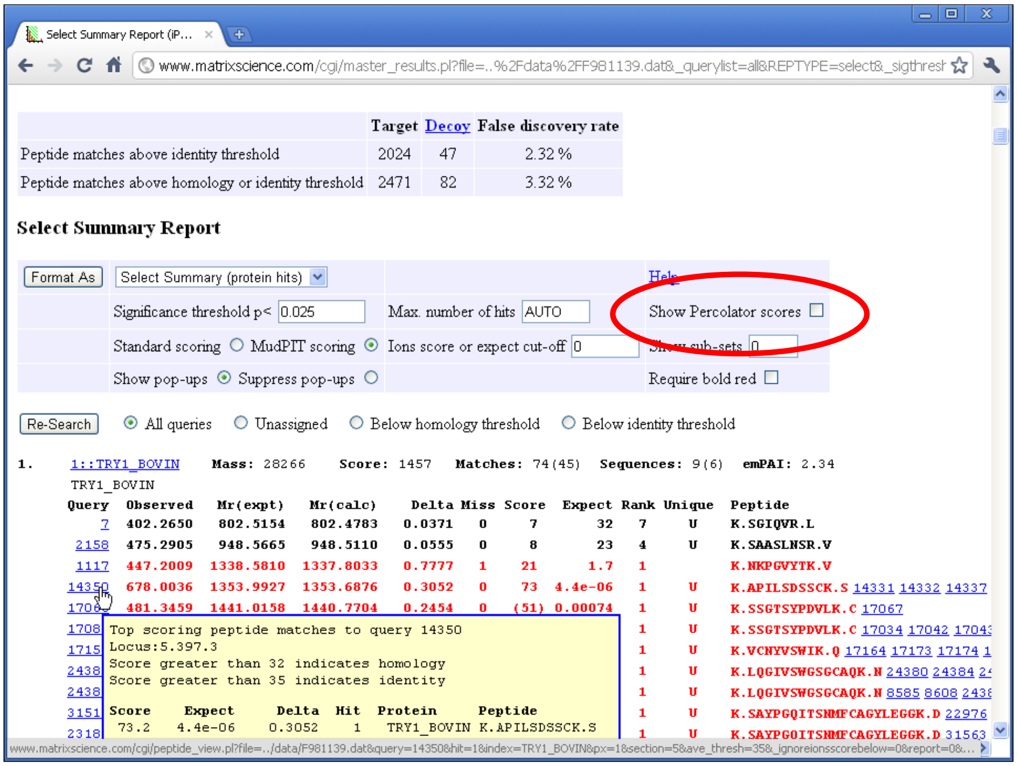
Re-Search ● All queries ○ Unassigned ○ Below homology threshold ○ Below identity threshold

```
1.     1::TRY1_BOVIN     Mass: 28266     Score: 1457     Matches: 74(45)    Sequences: 9(6)    emPAI: 2.34
       TRY1_BOVIN
      Query  Observed    Mr(expt)    Mr(calc)     Delta Miss Score  Expect Rank Unique   Peptide
          7  402.2650    802.5154    802.4783    0.0371    0     7      32    7     U     K.SGIQVR.L
       2158  475.2905    948.5665    948.5110    0.0555    0     8      23    4     U     K.SAASLNSR.V
       1117  447.2009   1338.5810   1337.8033    0.7777    1    21      1.7   1           K.NKPGVYTK.V
      14350  678.0036   1353.9927   1353.6876    0.3052    0    73  4.4e-06   1     U     K.APILSDSSCK.S 14331 14332 14337
       1706  481.3459   1441.0158   1440.7704    0.2454    0  (51)           1     U     K.SSGTSYPDVLK.C 17067
       1708                                                                 1     U     K.SSGTSYPDVLK.C 17034 17042 1704
       1715                                                                 1     U     K.VCNYVSWIK.Q 17164 17173 17174 1
       2438                                                                 1     U     K.LQGIVSWGSGCAQK.N 24380 24384 2
       2438                                                                 1     U     K.LQGIVSWGSGCAQK.N 8585 8608 243
       3151                                                                 1     U     K.SAYPGQITSNMFCAGYLEGGK.D 22976
       2318                                                                 1     U     K.SAYPGQITSNMFCAGYLEGGK.D 31563
```

```
Top scoring peptide matches to query 14350
Locus:5.397.3
Score greater than 32 indicates homology
Score greater than 35 indicates identity

Score     Expect     Delta   Hit   Protein     Peptide
 73.2    4.4e-06    0.3052    1    TRY1_BOVIN  K.APILSDSSCK.S
```

We also display an expect value. The expect value reports the quality of an individual match using a single number. It is the number of times in the search we could expect to get a match with this score or higher by chance. Low values are good, and an expect value of 1 or more is likely to be a random match. This match, with the score of 73, is a very strong match, and has a 1 in 200,000 chance of being a random match. The expect value is not new information; it could be calculated from the score and the identity threshold. It is a cleaner way of presenting the quality of a match because you don't have to keep looking at the difference between the score and the threshold



	Target	Decoy	False discovery rate
Peptide matches above identity threshold	2024	47	2.32 %
Peptide matches above homology or identity threshold	2471	82	3.32 %

Select Summary Report

Significance threshold p< 0.025
Max. number of hits AUTO
Show Percolator scores ☐
Standard scoring ○ MudPIT scoring ●
Ions score or expect cut-off 0
Show sub-sets 0
Show pop-ups ● Suppress pop-ups ○
Require bold red ☐

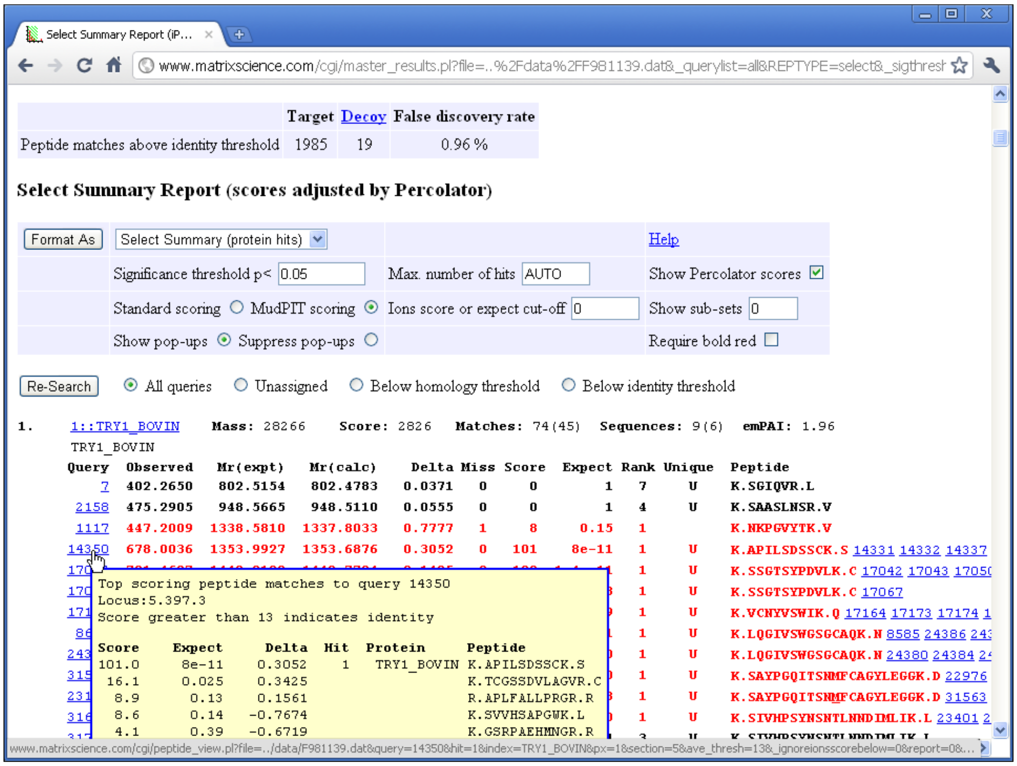
● All queries ○ Unassigned ○ Below homology threshold ○ Below identity threshold

```
1.      1::TRY1_BOVIN     Mass: 28266     Score: 1457    Matches: 74(45)   Sequences: 9(6)   emPAI: 2.34
        TRY1_BOVIN
       Query  Observed    Mr(expt)   Mr(calc)     Delta Miss Score  Expect Rank Unique   Peptide
           7  402.2650    802.5154   802.4783    0.0371    0     7      32    7     U    K.SGIQVR.L
        2158  475.2905    948.5665   948.5110    0.0555    0     8      23    4     U    K.SAASLNSR.V
        1117  447.2009   1338.5810  1337.8033    0.7777    1    21     1.7    1          K.NKPGVYTK.V
       14350  678.0036   1353.9927  1353.6876    0.3052    0    73 4.4e-06    1     U    K.APILSDSSCK.S 14331 14332 14337
        1706  481.3459   1441.0158  1440.7704    0.2454    0  (51) 0.00074    1     U    K.SSGTSYPDVLK.C 17067
        1708                                                                1     U    K.SSGTSYPDVLK.C 17034 17042 1704
        1715                                                                1     U    K.VCNYVSWIK.Q 17164 17173 17174 1
        2438                                                                1     U    K.LQGIVSWGSGCAQK.N 24380 24384 24
        2438                                                                1     U    K.LQGIVSWGSGCAQK.N 8585 8608 2438
        3151                                                                1     U    K.SAYPGQITSNMFCAGYLEGGK.D 22976
        2318                                                                1     U    K.SAYPGQITSNMFCAGYLEGGK.D 31563
```

```
Top scoring peptide matches to query 14350
Locus:5.397.3
Score greater than 32 indicates homology
Score greater than 35 indicates identity

Score     Expect      Delta   Hit   Protein     Peptide
 73.2    4.4e-06     0.3052     1   TRY1_BOVIN K.APILSDSSCK.S
```

In Mascot 2.3, for large searches that include an automatic decoy search, we also have the option to process the results through Percolator. This was described in some detail in last year’s meeting and the talk can be found on our web site.



	Target	Decoy	False discovery rate
Peptide matches above identity threshold	1985	19	0.96 %

Select Summary Report (scores adjusted by Percolator)

Format As: Select Summary (protein hits)

Significance threshold p< 0.05 · Max. number of hits AUTO · Show Percolator scores ☑

Standard scoring ○ MudPIT scoring ● · Ions score or expect cut-off 0 · Show sub-sets 0

Show pop-ups ● Suppress pop-ups ○ · Require bold red ☐

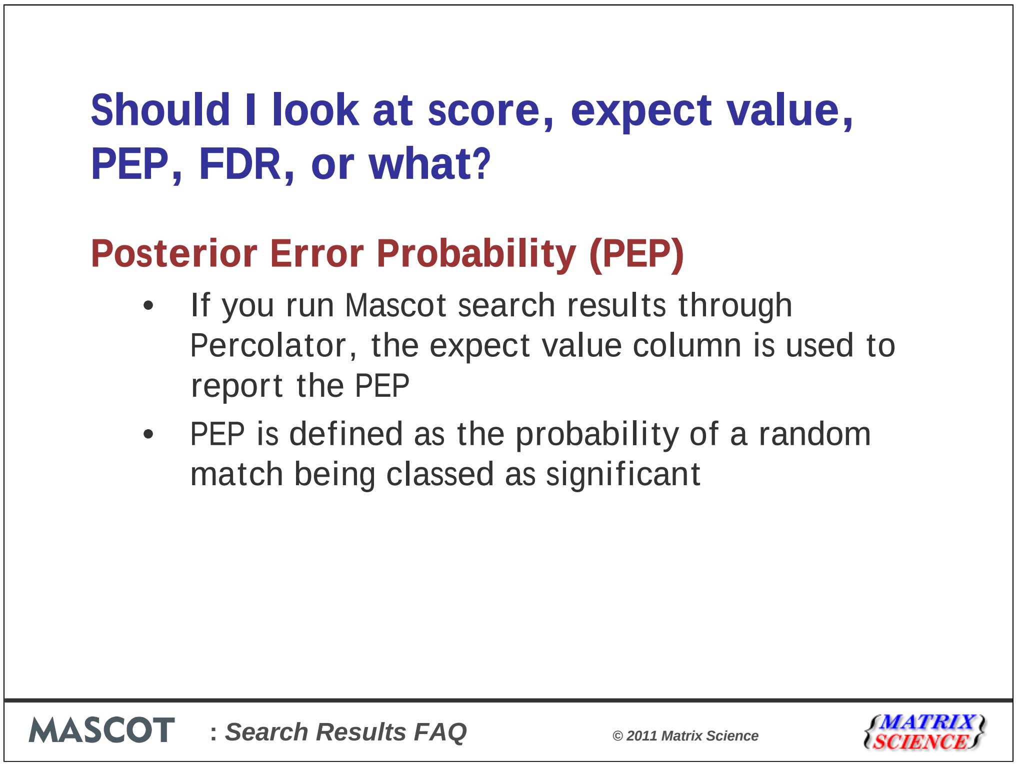
● All queries ○ Unassigned ○ Below homology threshold ○ Below identity threshold

```
1.      1::TRY1_BOVIN     Mass: 28266     Score: 2826    Matches: 74(45)   Sequences: 9(6)   emPAI: 1.96
        TRY1_BOVIN
       Query  Observed    Mr(expt)   Mr(calc)    Delta Miss Score   Expect Rank Unique    Peptide
           7  402.2650    802.5154   802.4783   0.0371    0     0        1    7     U      K.SGIQVR.L
        2158  475.2905    948.5665   948.5110   0.0555    0     0        1    4     U      K.SAASLNSR.V
        1117  447.2009   1338.5810  1337.8033   0.7777    1     8     0.15    1            K.NKPGVYTK.V
       14350  678.0036   1353.9927  1353.6876   0.3052    0   101    8e-11    1     U      K.APILSDSSCK.S 14331 14332 14337
       170                                                                  1     U      K.SSGTSYPDVLK.C 17042 17043 1705
       170                                                                  1     U      K.SSGTSYPDVLK.C 17067
       171                                                                  1     U      K.VCNYVSWIK.Q 17164 17173 17174 1
        86                                                                  1     U      K.LQGIVSWGSGCAQK.N 8585 24386 24
       243                                                                  1     U      K.LQGIVSWGSGCAQK.N 24380 24384 2
       315                                                                  1     U      K.SAYPGQITSNMFCAGYLEGGK.D 22976
       231                                                                  1     U      K.SAYPGQITSNMFCAGYLEGGK.D 31563
       316                                                                  1     U      K.SIVHPSYNSNTLNNDIMLIK.L 23401 2
```

```
Top scoring peptide matches to query 14350
Locus:5.397.3
Score greater than 13 indicates identity

Score     Expect     Delta   Hit   Protein      Peptide
101.0      8e-11    0.3052     1    TRY1_BOVIN  K.APILSDSSCK.S
 16.1      0.025    0.3425                      K.TCGSSDVLAGVR.C
  8.9       0.13    0.1561                      R.APLFALLPRGR.R
  8.6       0.14   -0.7674                      K.SVVHSAPGWK.L
  4.1       0.39   -0.6719                      K.GSRPAEHMNGR.R
```

If we check the box and choose ‘Format As’, Percolator calculates something called a Posterior Error Probability (PEP) for each match, and this is displayed in the expect value column. To avoid breaking software that parses Mascot results and expects to get a score and a threshold for each match, we calculate a fake score from the PEP relative to a fixed score threshold of 13. The default FDR for results after running Percolator is initially 1%



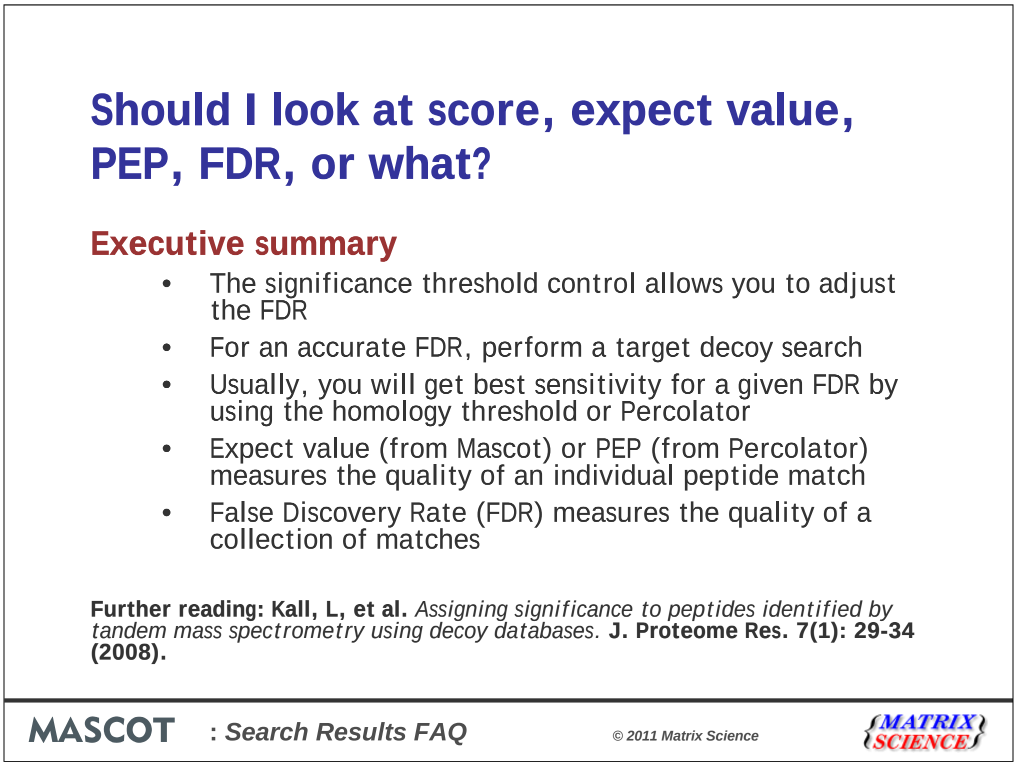
Should I look at score, expect value, PEP, FDR, or what?

Posterior Error Probability (PEP)

- If you run Mascot search results through Percolator, the expect value column is used to report the PEP
- PEP is defined as the probability of a random match being classed as significant

MASCOT : *Search Results FAQ*

What is Posterior Error Probability? It is defined as the probability of a random match being classed as significant. This is not the same as an expect value, which is the number of times in the search we could expect to get a random match classed as significant, but there is little difference between the two for values < 0.1



Should I look at score, expect value, PEP, FDR, or what?

Executive summary

- The significance threshold control allows you to adjust the FDR
- For an accurate FDR, perform a target decoy search
- Usually, you will get best sensitivity for a given FDR by using the homology threshold or Percolator
- Expect value (from Mascot) or PEP (from Percolator) measures the quality of an individual peptide match
- False Discovery Rate (FDR) measures the quality of a collection of matches

Further reading: Kall, L, et al. *Assigning significance to peptides identified by tandem mass spectrometry using decoy databases.* **J. Proteome Res. 7(1): 29-34 (2008).**

To summarise.