

New features in Mascot Server 2.5

MASCOT : *Mascot Server 2.5*

© 2014 Matrix Science

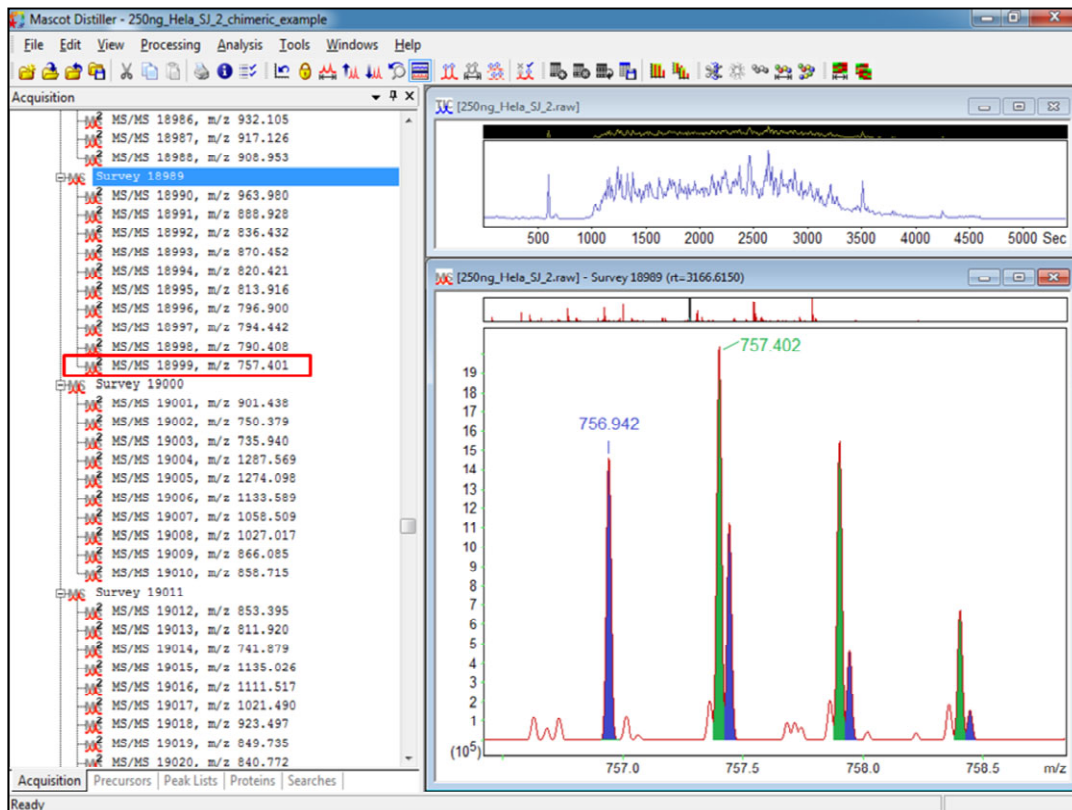


New features in Mascot Server 2.5

- Search engine
- Reports
- Administration (databases, modifications)
- Mascot Parser
- Mascot Daemon

We've just released Mascot Server 2.5 and in this talk I'll be covering some of the new features.

There are changes in the search engine and reports. There are also improvements to the administration of databases and some significant enhancements to the way modifications are handled. I'll also be covering changes in Mascot Parser and Mascot Daemon



Starting off with the search engine, this first enhancement makes better use of high resolution data. I'm going to illustrate this initially with screen shots from Mascot Distiller using a Q Exactive dataset (250 ng Hela digest over 90 min run) kindly provided by Thermo. However, this process will work with high resolution data from any instrument. You can see a survey scan with 10 ms-ms scans following it. I'm just going to look at this one ms-ms scan, with a memorable id of 18,999 and an m/z of 757.401. One of the real benefits of Mascot Distiller is that it can go back and look at the survey scan and re-determine the charge state and m/z for each ms-ms scan. If we zoom in on the survey scan at about the 757 mass, we can see what looks like 2, or maybe even 3 overlapping peak clusters – this one at 756.942 and this one at 757.402. We, as intelligent humans, and of course Distiller can't be sure at this stage, but it's quite likely that the isolation window on the Q Exactive will have let both peptides through for fragmentation, and we stand a chance of identifying both peptides. If we process this data and submit a search to Mascot Server 2.5, then

Chimeric spectra

```
BEGIN IONS
TITLE=15397: Scan 18999 (rt=3168.16)
PEPMASS=756.94206 1583103.2 2+
PEPMASS=757.40156 2243692 2+
CHARGE=2+
SCANS=18999
RAWSCANS=sn18999
RTINSECONDS=3168.1628
105.05717      5843.5938
. . .
END IONS
```

MASCOT : Mascot Server 2.5

© 2014 Matrix Science



...this is mgf file that Distiller sends to the server. Note that it has 2 PEPMASS lines – one for each precursor match it has found.

Chimeric spectra

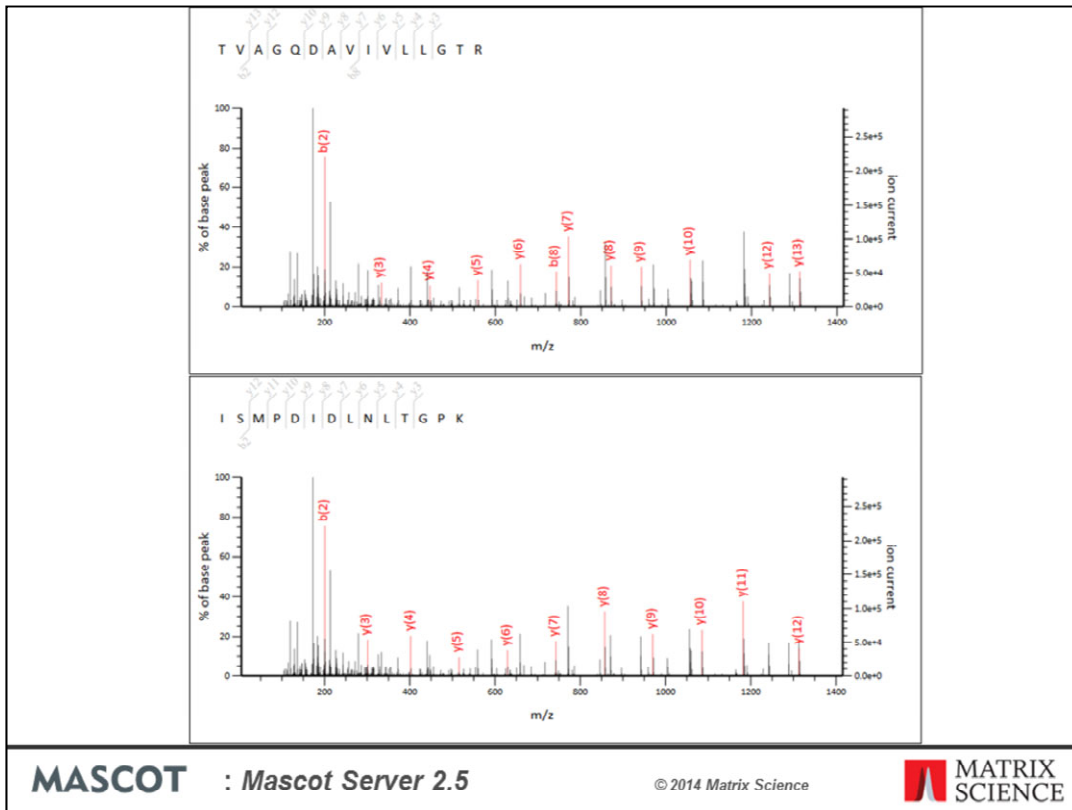
Query Dupes	Observed	Mr(expt)	Mr(calc)	ppm	M Score	Expect	Rank	U	Peptide
<input checked="" type="checkbox"/> 20796	756.94206	1511.86957	1511.86716		1.59	0.57	4.1e-006	▼1	U K.TVAGQDAVIVLLGTR.N
15397: Scan 18999 (rt=3168.16) [D:\data_Distiller test data\11537 Score > 20 indicates identity Score > 16 indicates homology No other peptide matches in query									
<input checked="" type="checkbox"/> 20797	757.40156	1512.78857	1512.78577		1.85	0.34	0.00066	▼1	U K.ISMPDIDLNLTPK.I
15397: Scan 18999 (rt=3168.16) [D:\data_Distiller test data\11537 Score > 23 indicates identity Score > 15 indicates homology No other peptide matches in query									

MASCOT : Mascot Server 2.5

© 2014 Matrix Science



And in the results we can see two separate queries for the same scan, number 18999, and two quite different peptide matches, both with a significant score.



And if you compare the peptide views for each of the two spectra carefully, you can see that different peaks are being matched.

Chimeric spectra - example data set

- 21962 ms-ms spectra
- 42013 queries (i.e. average of 2 precursors)
- 9245 significant peptide matches
- 209 significant 'chimeric' matches
- 630 significant matches to less intense precursor
- 8381 significant peptide matches with ver 2.4

MASCOT : Mascot Server 2.5

© 2014 Matrix Science



Looking at some statistics for the data set as a whole, there were about 22 thousand ms-ms spectra. Distiller found an average of 2 possible precursor matches for each of these spectra, so we have double the number of queries. Searching this data file with tight tolerances gives us 9245 significant peptide matches. This includes 209 chimeric matches – that is two matches to the same ms-ms spectrum but with a different precursor. Also, perhaps more significantly, there were 630 cases where the most intense precursor from the MS spectrum failed to get a match, but the less intense precursor gave a significant match. Summing these two numbers pretty much accounts for the difference you'll see between Mascot 2.4 and Mascot 2.5

It's perhaps worth pointing out that using Mascot 2.4, you could get some of these 630 significant matches by using a wide precursor tolerance.

Specifying the data file as a URL

The screenshot shows the Mascot search interface. At the top, 'Peptide charge' is set to '2+' and 'Monoisotopic' is selected. Below this, there are two radio buttons: 'Data file' (unselected) and 'Data URL' (selected). The 'Data URL' option is followed by a text input field containing the path: 'ftp://PASS00138:IG4328f@ftp.peptideatlas.org/IEC6-NuclPD_0min-0mM-R2-1.mgf'. Below the input field, 'Data format' is set to 'Mascot generic' and 'Precursor' is set to 'm/z'.

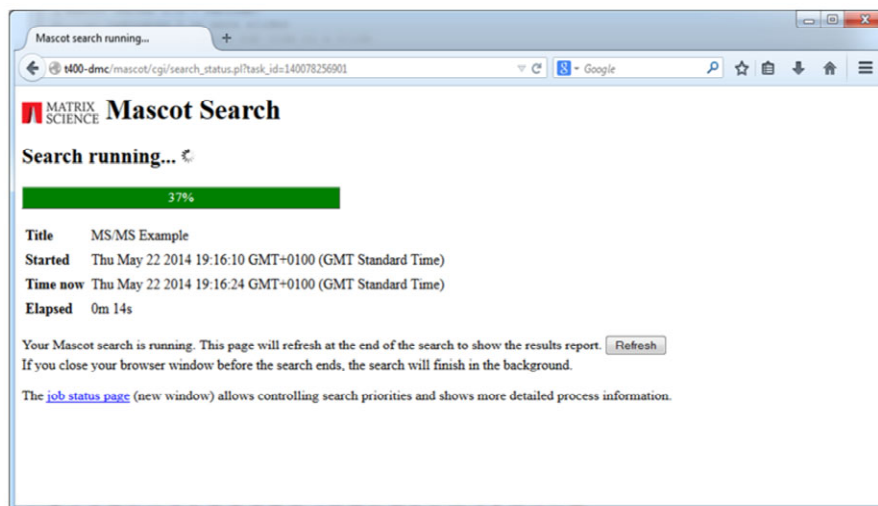
- Now possible to add “Search with Mascot” links from a web site
- This option has to be enabled in Mascot Security

This next new feature is mostly going to be of use to web site designers and people writing MS software. However, we’ve had quite a few requests for it over recent years.

Instead of using a local file and browsing to that file, you give a URL for the location of the file. The file needs to be available via http, ftp or it could be a local file specified using the file:// URL.

Amongst other uses, it will enable developers to put links on web pages which allow people to just submit a Mascot search without having to download and then upload a data file.

Asynchronous Search submission!



MASCOT : Mascot Server 2.5

© 2014 Matrix Science



One thing you'll notice straight away if you submit searches through a web browser is that there are different progress reports. No more dots...

The reason for this change was that using IIS under Windows it was getting harder and harder to stop searches being killed by the web server when the browser had a slight glitch and got disconnected. This asynchronous technique, gives us much more control over this, so long searches should carry on to completion more reliably. There's also a link to the job status page here so that you can easily change the priority of the search, or terminate it if you wish.

Report improvements

- **Modification statistics**

▶ Score distribution

▼ Modification statistics

Modification	Site	Above thr.	ET	Total matches
Oxidation	M	13	0	13
Acetyl	N-term	8	0	8
Carbamidomethyl	C	7	0	7
Non-specific cleavage	-	0	5	5
Carbamidomethyl	N-term	0	3	3
Carbamidomethyl	K	0	2	2
Ethanedithiol	S	0	2	2
Amidine	N-term	0	2	2
Deamidated	N	0	1	1
Phe->Val	F	0	1	1
CAMthiopropionyl	K	0	1	1
Gly->Val	G	0	1	1
Glu->Lys	E	0	1	1
Succinyl	N-term	0	1	1
Methyl	E	0	1	1
Leu->His	L	0	1	1
Amino	Y	0	1	1
Iodo	Y	0	1	1
Carboxymethyl	N-term	0	1	1
...

MASCOT : Mascot Server 2.5

© 2014 Matrix Science



There's a widget at the top of the report that you can click to display the count of modifications found in the search.

When you see something of interest, you can of course use the search facility in the report to find all the instances of that modification.

This report is very useful with an error tolerant search as it will quickly show you which are the most common found modifications. This example is actually a very small search, but it does show up a likely problem with the sample preparation. The 5 additional matches found for N term and lysine carbamidomethyl are almost certainly the result of over carbamidomethylation. You probably want to change the protocol, but in the meantime it may be best to add these as variable mods to your searches. However, you'll notice that carbamidomethyl (C) is in the short list of mods, but the N term and lysine mods are missing. I'll illustrate how to change this using the improved configuration editor in a later slide.

Report improvements

- For iTRAQ, TMT etc. report ratio p-values
- Added pI and protein coverage as columns
- Show the quantitation component for all precursor methods
- Normalise to selected protein accessions or peptide sequences

MASCOT : Mascot Server 2.5

© 2014 Matrix Science

 **MATRIX
SCIENCE**

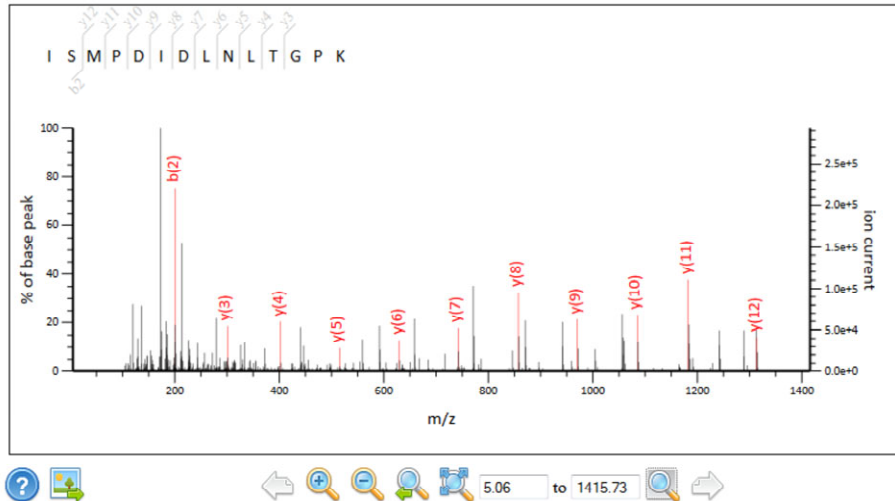
For iTRAQ, TMT etc. we now report p-values for quantitation ratios, and not just whether significantly different from 1

We have added Isoelectric point and protein coverage as optional columns in the report builder

We used to just show the quantitation component for 15N and other metabolic labelling methods, but we now show it for all precursor methods, such as SILAC. Of course, to get quantitation values, this still needs to be done in Mascot Distiller

For Reporter quantitation methods such as iTRAQ and TMT, you can now specify a protein accession or peptide sequence to normalise against.

New peptide view reports



MASCOT : Mascot Server 2.5

© 2014 Matrix Science

MATRIX SCIENCE

If the report is viewed in most modern web browsers, the spectrum display is an interactive SVG graphic that supports zoom and pan.

Our thanks go to Juri Rappsilber and Jimi-Carlo Bukowski-Wills from the University of Edinburgh for allowing us to incorporate this viewer in to Mascot (<http://spectrumviewer.org/>).

The Help button on the left gives a short summary of how to use the new display. The icon just to the right of the help button is for use in exporting the image as an SVG file. This is much superior to the bitmaps we have had in the previous versions of Mascot and makes putting high quality images into publications and papers really simple.

Zooming in is easy and interactive by dragging out a region in the X axis. Scroll using the mouse wheel if you have one, or by using these arrows.

Hover the mouse over a cleavage point in the peptide sequence to highlight the corresponding peaks in the spectrum and vice versa. Drag between two peaks to display mass differences. The controls have tooltip help.

Exporting the new graphic

- The export button saves an SVG file - Scalable Vector Graphics
- When it's expanded, maybe on a large poster, the quality remains high
- Most graphics programs can import SVG
- For use with Word, download the free Inkscape software and save as an Enhanced Metafile.

The export button saves an SVG file – this stands for Scalable Vector Graphics. The difference between this and say a JPEG file is that as you expand the image, for example on a poster, the quality remains high and doesn't become blocky.

If you have software like Adobe Illustrator, then there's no problem with loading an SVG file.

If you are using Microsoft Word, then you'll need something like the excellent Inkscape freeware. Load the SVG into Inkscape and then Save As an Enhanced Metafile. You can then paste it into Microsoft Word or Powerpoint and it will still be scalable in that application.

Fine grain modification control

- Changes to defaults now kept separately
- Easier to update to latest unimod.xml:
- With Mascot security, groups can hide modifications from another group
- Specify modifications used for error tolerant search
- Easier to change list of displayed mods.

In previous versions of Mascot, any custom modifications or changes to what were displayed in the short list in the search form were added to the default unimod.xml file. That has meant that for each new Mascot release, when a new unimod.xml file was supplied, all your customisations used to get lost. From 2.5 onwards, any changes you make will be saved to separate files. The bad news is that you are going to have to suffer this one more time as the update to 2.5 will lose your custom changes. The good news is that's the last time you'll need to do it as all configurations changes will be stored in a separate usermods.xml file. So, when you next update the unimod.xml file, either as part of a future Mascot update, or by downloading it, your own customisations will be merged into the master file. We've also made it easier to get the latest unimod.xml file too – just click on the 'Check Unimod' button in the editor.

If you have Mascot Security enabled, one group can have a private modification that another group can't see – something we've been asked for more times than you might expect.

One of the other reasons for these improvements is that it's now possible to specify which modifications are used for an error tolerant search. It's often annoyed people that with an error tolerant search, Mascot finds a good match to a peptide with, say an ICPL modification but when there's not even any ICPL in the lab. Now, we've always argued that this is potentially a good thing because it means that something is causing a delta of 115 daltons – it's just that you need to go and find out what it is. We've given in to the pressure to change this, and I'll show you how easy it is to configure this on the next slide.

Another useful improvement is the ability to change what's shown in the short list of

modifications on the search form.

Mascot Configuration: Modifications Administrator Logout

Displaying 48/944

Visibility:

- Short list
- Long list
- Mixed
- Not listed

Error tolerant:

- Yes
- No
- Mixed

Classifications: clear

-
- Post-translational
- Co-translational
- Pre-translational
- Chemical derivative
- Artefact

Source:

- Unimod
- Edited Unimod
- Local

Apply to selected: 0

- Include in short list
- Include in long list
- Include in error tolerant
- Exclude from error tolerant

Modifications

Title	Monoisotopic	Average	Composition	Source	Visibility	Err Tol
<input type="checkbox"/> Acetyl	42.010565	42.0367	C(2) H(2) O	edited	mixed	yes
<input type="checkbox"/> Amidated	-0.984016	-0.9848	H N O(-1)	Unimod	short	yes
<input type="checkbox"/> Ammonia-loss	-17.026549	-17.0305	H(-3) N(-1)	Unimod	mixed	yes
<input type="checkbox"/> Biotin	226.077598	226.2954	H(14) C(10) N(2) O(2) S	Unimod	short	yes
<input checked="" type="checkbox"/> Carbamidomethyl	57.021464	57.0513	H(3) C(2) N O	Unimod	mixed	yes
<input type="checkbox"/> Carbamyl	43.005814	43.0247	H C N O	Unimod	mixed	yes
<input type="checkbox"/> Carboxymethyl	58.005479	58.0361	H(2) C(2) O(2)	Unimod	mixed	yes
<input type="checkbox"/> Cation:Na	21.981943	21.9818	H(-1) Na	Unimod	short	yes
<input type="checkbox"/> cysTMT6plex	304.177202	304.3962	H(25) C(10) 13C(4) N(2) 15N O(2) S	Unimod	short	yes
<input type="checkbox"/> Deamidated	0.984016	0.9848	H(-1) N(-1) O	Unimod	mixed	yes
<input type="checkbox"/> Dehydrated	-18.010565	-18.0153	H(-2) O(-1)	Unimod	mixed	yes
<input type="checkbox"/> Dehydro	-1.007825	-1.0079	H(-1)	Unimod	short	yes
<input type="checkbox"/> DiLeu4plex	145.132163	145.2229	H(13) 2H(2) C(8) N 18O	Unimod	short	yes
<input type="checkbox"/> Dioxidation	31.989829	31.9988	O(2)	Unimod	mixed	yes
<input type="checkbox"/> Ethanolyl	44.026215	44.0526	H(4) C(2) O	Unimod	mixed	yes
<input type="checkbox"/> ExacTagAmine	1046.347854	1046.8285	H(52) C(25) 13C(12) N(8) 15N(6) O(19) S	Unimod	short	yes
<input type="checkbox"/> ExacTagThiol	972.365219	972.7268	H(50) C(23) 13C(12) N(8) 15N(6) O(18)	Unimod	short	yes
<input type="checkbox"/> Formyl	27.994915	28.0101	C O	Unimod	mixed	yes
<input type="checkbox"/> Gln->pyro-Glu	-17.026549	-17.0305	H(-3) N(-1)	Unimod	short	yes
<input type="checkbox"/> Glu->pyro-Glu	-18.010565	-18.0153	H(-2) O(-1)	Unimod	short	yes

Page 1/3 Go to page Ac to GI << >> Page size 20

Add new modification Main menu Check Unimod

There's a lot of new functionality in the configuration editor to enable this. For example, you can filter modifications to just show chemical derivatives, or in this case I'm filtering to just show what's on the short list of modifications you see in the search form. You'll notice that I've selected the 'short list' check box here and also the 'mixed' check box. Mixed means that some specificities are shown in the short list, and some in the complete long list. So for example, carbamidomethyl is shown as mixed. You may remember from an earlier slide, I promised to show you how we could add N term and lysine carbamidomethyl to the short list of mods. If I click on the carbamidomethyl link

<http://1400-dmc/mascot/x-cgi/ms-config.exe?u=1401800843> Mascot configuration

Edit Modification : Carbamidomethyl

Administrator [Logout](#)

Name

Title

Fullname

[Delta](#)
[Specificity](#)
[Ignore Masses](#)
[Misc](#)
[References](#)

Specificity

Specificity	Site	S	Position	Anywhere	Copy	Delete	Show Details
Specificity	Site	E	Position	Anywhere	Copy	Delete	Show Details
Specificity	Site	D	Position	Anywhere	Copy	Delete	Show Details
Specificity	Site	H	Position	Anywhere	Copy	Delete	Show Details
Specificity	Site	N-term	Position	Any N-term	Copy	Delete	Show Details
Specificity	Site	K	Position	Anywhere	Copy	Delete	Show Details
Specificity	Site	C	Position	Anywhere	Copy	Delete	Show Details
Specificity	Site	T	Position	Anywhere	Copy	Delete	Show Details
Specificity	Site	Y	Position	Anywhere	Copy	Delete	Show Details

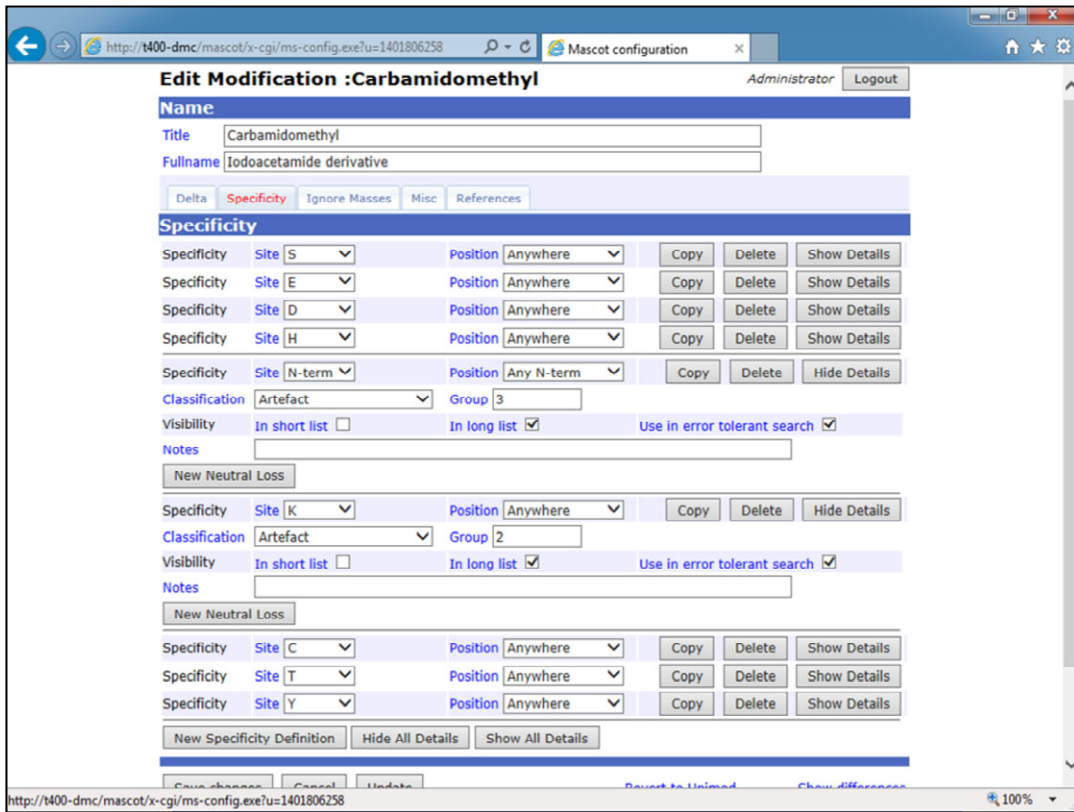
[New Specificity Definition](#)
[Show All Details](#)

[Save changes](#)
[Cancel](#)
[Update](#)
[Revert to Unimod](#)
[Show differences](#)

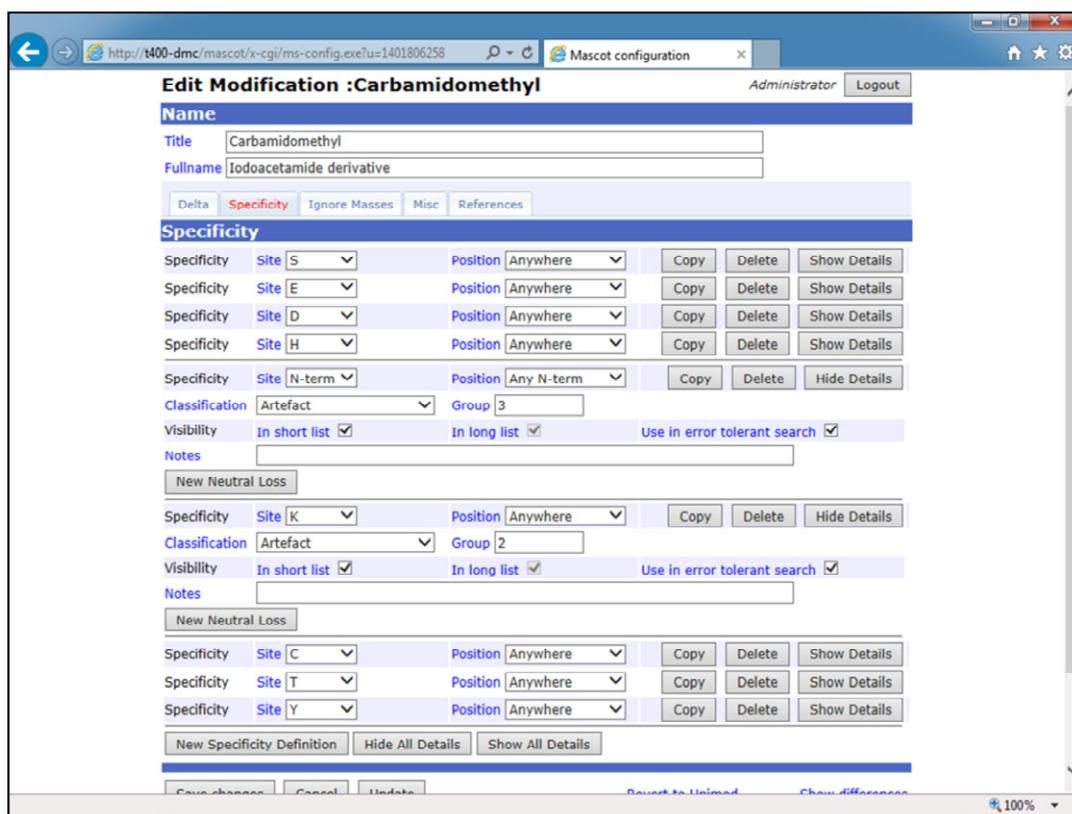
Help Window

100%

And then on the specificity tab



And then if I Show Details for the N-term and lysine specificities, you'll see that I can easily add these to the short list.



And of course these changes will now be preserved when you update to the latest unimod.xml or install a new version of Mascot.

Database Manager improvements

- **Easier to add a SwissProt type database without a reference file**
- **New overview table of update schedules**
- **Improved proxy handling**
- **FASTA and reference file uploading through the web browser; no need to have access to the Mascot Server hard disk.**

MASCOT : Mascot Server 2.5

© 2014 Matrix Science



We've been slowly making minor improvements to database manager based on feedback from many of our users.

It's now easier to add a database that uses the SwissProt fasta format without having a corresponding reference file.

It wasn't easy to see a list of all the updates scheduled before, but there's a new table for this now

We've improved proxy handling, not just in database manager but for other utilities and for the protein view report

Also, it's now possible to upload a new fasta file without having direct access to the disk on the Mascot server.

Miscellaneous Mascot Server changes

- Help tutorial added
- mzML processing much faster
- Number of variable mods greater than 9
- Search progress displayed differently
- Improved memory handling
- Performance issue in cluster mode
- ITOL in ppm
- And more . . .

Mascot Parser

- **Quantitation routines**
 - emPAI
 - Reporter
- **Ability to open Distiller files**
- **Add classes to facilitate communication with a remote Mascot server via HTTP**
- **All the above help when writing software that uses Mascot Server**

This slide is also only of interest to developers, but it's quite important.

We've added a lot of support for quantitation to Mascot Parser and also the ability to open Distiller files. If you are working on software that integrates with Mascot Server, I'd strongly recommend that you download the latest version of Mascot Parser and look at the new additions.

Mascot Daemon - major upgrade

- **64 bit and 32 bit versions now available**
- **Fully utilises all cores with Distiller processing**
- **Better support for multi-byte language filenames.**
- **Improved installer**
- **Can now run Daemon as a 'normal user' - admin rights only needed for installation.**

MASCOT : Mascot Server 2.5

© 2014 Matrix Science



We've actually re-written a large part of Mascot Daemon for this release. You may not notice this, because it actually looks very similar to the previous version. One of the reasons for the significant changes under the hood was to give better integration with Distiller. To do this, we had to have a 64 bit version so that more memory was available to enable the use of more processor cores. With the 32 bit version in 2.4 and earlier, there were memory constraints and you needed to tweak a Windows registry value to be able to use a few cores, but there were still severe limitations. When you run Daemon 2.5 on a 64 bit system, you'll find that Distiller uses as many cores as are available for peak picking and quantitating raw data.

Previous versions used the Microsoft Access database engine by default. With the move to 64 bit, this was no longer quite so easy, so for a new clean installation, we use an integrated database engine. There are tools to migrate the data from the old taskDB.mdb file to the new VistaDB engine. Full details in the help.

Part of the major re-write has enabled us to provide better support for far eastern language filenames such as Chinese and Japanese.

The installer has also been re-written and should be much smoother. It's now also much easier to run it as a normal user and you should only need to be administrator to actually install the software.

Mascot Daemon

- **New integrated database engine instead of Microsoft Access**
- **Still provide support for many other databases such as SQL Server or Oracle**
- **Multiple instances on single PC using RDP**
- **Search parameters in database option**
- **Time-out problems for large searches fixed**
- **Insight integration**

MASCOT : Mascot Server 2.5

© 2014 Matrix Science



Previous versions used the Microsoft Access database engine by default. With the move to 64 bit, this was no longer quite so easy, so for a new clean installation, we use an integrated database engine. There are tools to migrate the data from the old taskDB.mdb file to the new integrated database engine. Full details in the help.

It's still possible to use any of the databases that were supported in previous versions such as Microsoft SQL Server, Oracle or MySQL. We now also provide support for Postgres.

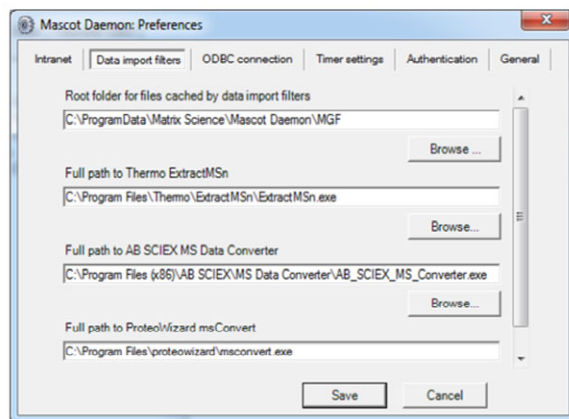
It's now possible to run multiple instances of Daemon as different users on single server PC and for the different users to connect to the server using remote desktop.

There's now an option to save search parameters in the database rather than as files. If Daemon installations on multiple PCs share the same database on a network drive, this is a much more convenient way to share search parameters.

Some issues with time out problems for large searches have also been addressed and this should be more robust.

And, as we've already seen, there is integration with Mascot Insight.

Mascot Daemon - Import filters



MASCOT : Mascot Server 2.5

© 2014 Matrix Science



We have added two new import filters – the AB SCIEX data converter and ProteoWizard. If you've been used to using Mascot.DLL in the past, you'll notice that this is no longer an option because it doesn't support newer files and is no longer being updated by AB SCIEX. We strongly recommend switching to the new AB SCIEX MS Data Converter as this also does a much better job of peak picking. The MS Data Converter also supports TOF-TOF data as well as the Analyst wiff files.

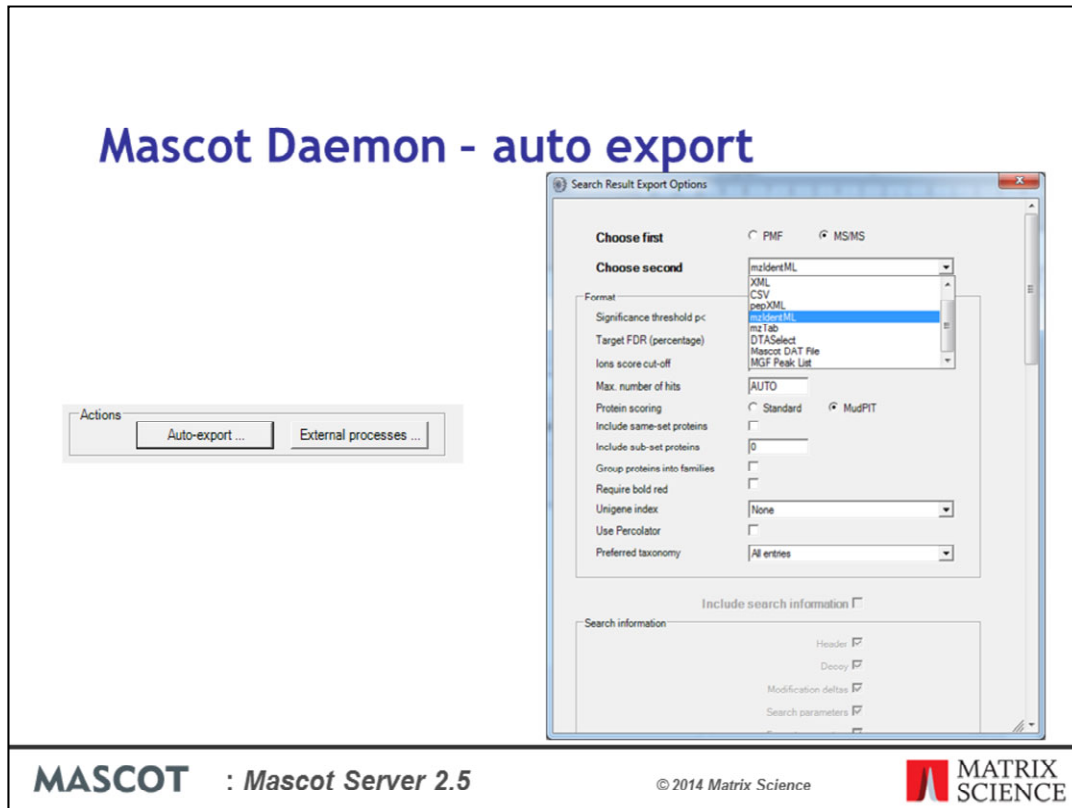
The ProteoWizard msConvert application will convert most types of RAW file to either MGF or mzML which can be used by Mascot. You'll find that you get better results using Mascot Distiller.

You'll find full details in the Daemon help – just press F1 – and follow the links for downloading.

We are still supporting ExtractMSn.exe, but Thermo have stopped supporting this and it doesn't work with more recent instruments and versions of Xcalibur. If you were using successfully in the past, it should continue to function with the update to Mascot 2.5

In Mascot Daemon 2.5 it's also possible to add your own import filters by editing a configuration file. Once again, details in the help.

Mascot Daemon - auto export



In previous versions of Mascot Daemon, it was possible to automatically export the search results to a file for each search, but this required some additional scripts which needed editing and it was quite hard to set up. We've now included a button on the task editor tab and this brings up window where can select the required options. There are a lot of export formats to choose from and another requested feature is the ability to select a target FDR. Obviously, for this to work, you need to select the option for a decoy search in your search parameters. The export files are saved in the same directory as the mgf and project files.

This is the same list of export formats that you'll see in Mascot Server. We've added support the mzTab export format from the HUPO Proteomics Standards Initiative. This is a fairly simple format that supports quantitation but note that the standard hasn't yet been finalised. We are supporting the latest release candidate which is release 5, from December last year. We output quantitation data from the quantitation techniques already supported in Mascot Server like iTRAQ , TMT and multiplex.

It's now also possible to export percolator p and q values and also the normalisation factors used in quantitation

New features in Mascot Server 2.5

- Search engine
- Reports
- Administration (databases, modifications)
- Miscellaneous
- Mascot Parser
- Mascot Daemon
- New web page design!

MASCOT : Mascot Server 2.5

© 2014 Matrix Science

 **MATRIX
SCIENCE**

I've outlined some of the changes in Mascot Server 2.5, but inevitably missed out many small enhancements and bug fixes that may be very important for some of you.

I've shown support for chimeric spectra, a new spectrum viewer and quite a few changes to the administration tools. We've also seen some big changes to Mascot Daemon. Oh, and finally, you may notice that our new web page design and logo is included in Mascot Server 2.5