

Peak picking for profit and pleasure...

MASCOT : *Peak Picking*

© 2013 Matrix Science



For the next half hour I'm going to be describing some of the different options for peak peaking. The profit is with getting better protein ID or quantitation, but to be totally honest, the pleasure really only comes when it's all done and dusted and running smoothly.

From raw data to peak list

- Mascot Server needs 'text' peak lists and doesn't accept raw, binary data
- Raw data files mostly contain profile data, but Mascot needs 'centroided' peak lists
- Raw data files may not contain accurate precursor masses or charges
- This talk describes your options for converting the raw data to peak lists.

Mascot Server needs 'text' peak lists and doesn't accept raw, binary data

Raw data files mostly contain profile data, so there will be 10s or even hundreds of data points for each peak, but Mascot Server needs just one monoisotopic peak mass and intensity value.

Raw data files may not contain accurate precursor masses or charges because these values need to be determined very quickly by the instrument control software while acquisition is taking place. For some data systems, the charge state is not saved at all.

This talk describes your options for converting the raw data to peak lists that can be searched by Mascot Server.

Factors affecting software choice

- Quality of the peak list
- Format of the peak list (e.g. MGF, mzML)
- Support for automation
- Are quantitation data required?
- Price
- Ease of use
- Integration with other pieces of software

MASCOT : *Peak Picking*

© 2013 Matrix Science



As with many things, there is mostly no perfect piece of software that will do all you need, and it's normally a case of finding something that does a good enough job.

The quality of the peak list will affect the number of matches you get from Mascot, and you should ideally compare different software and different parameters with your own data before making a choice. A good measure might be the number of significant peptide matches for a known standard sample at, say 1% FDR.

I'll describe the different peak list file formats that are accepted by Mascot. Most software only offers one or two output formats.

Being able to automate the production of the peak lists and submission of the searches may be important to you.

If you are performing quantitative experiments, there are some definite gotcha's that need to be considered.

Some software is free for everyone, some will be provided with your instrument, and some may cost additional dollars.

And as with all software, ease of use and integration with other packages may well limit your choices.

Quantitation Issues

- **MS-MS Based Quantitation - e.g. iTRAQ**
 - Important not to process the reporter ion region
- **MS1 Based Quantitation - e.g. SILAC**
 - Quality of the survey scans is important
 - If using Distiller, need to perform peak detection and the Mascot search from Distiller. Cannot use other tools for generating the peak list

For techniques where the quantitation information is in the ms-ms spectra, for example using isobaric tags to create reporter ions, it is important that the software you use does not attempt to treat these reporter as peptides. If it does, then it may, for example, treat the 117 reporter ion as a 13C peak for the 116 reporter ion, and remove it altogether.

For MS1 based quantitation, the quality of the survey scans is important. As well as choosing a suitable hi res method for collecting the data, you will probably need to save it in profile format rather than as centroided data.

If you are using Distiller for quantitation, you will need to perform peak detection and the Mascot search from Distiller. You cannot use other tools for generating the peak list for the protein ID.

Peak list format

Data file

Data format: Mascot generic

Instrument

Decoy

.raw : 250 Mb
.mgf : 300 Mb
.mzML : 2.5 Gb

MASCOT : Peak Picking

© 2013 Matrix Science

MATRIX
SCIENCE

Many of the software solutions for generating a peak list give you an option for the format of the peak list. As you will see from the Mascot search form, you have a number of options. I'll just describe why, given the choice, you don't want to choose most of these!

- The Sequest .dta file was one of the first formats for a single ms-ms spectrum. It has the precursor charge, precursor mass and a list of fragment m/z values but nothing else. The filename is typically used as a description and would often include the scan number or other identifier. A single ms-ms spectrum has limited value, so Mascot supports concatenated .dta files. However, if you do this, you lose the information that's included in the filename, so this isn't really a good idea.
- The .ASC files came from very early Finnigan instruments and are no longer in use.
- Micromass .PKL is very similar to multiple .dta files that have been concatenated except that the precursor mass is given as an m/z value, rather than as an MH+ value. Also, a precursor intensity is supplied, but similarly to the .dta files, there is no option for any additional data.
- The PerSeptive PKS and Sciex API III formats are not used by recent instruments.
- The Bruker .XML format isn't commonly used any more, only works for a single ms-ms spectrum and unfortunately doesn't contain the precursor mass, so is of very little value.
- mzData is an old format from the proteomics standard initiative and has been superseded by the mzML format. The advantage of this format is that it is possible to include lots of

additional data in a structured format. So, it would be possible for example to determine what instrument was used to collect the data and what methods were used during data collection and data processing. The downside is that the files tend to be much, much larger. If you need this additional information downstream for your workflow, then this is your best option.

However, in general, we recommend the mgf format which supports additional information such as retention time and scan numbers for each spectrum, but doesn't have the overhead of the mzML files. With a typical Thermo .raw file, the .mzML file was nearly a factor of 10 larger than the corresponding .mgf file.

Solutions

- **Matrix Science Solutions**
 - Mascot Daemon
 - Mascot Distiller
 - TS2Mascot
- **Instrument vendor software**
- **Free (sometimes open source) tools.**

MASCOT : *Peak Picking*

© 2013 Matrix Science



There are three different categories of software solutions, and I will go through these in turn.

Our offerings include Mascot Daemon, Mascot Distiller and TS2 Mascot

Each of the instrument manufacturers have their own software that integrates to varying degrees with Mascot. I'll outline each of these in turn, strictly in alphabetical order.

Finally, there are various other third party software offerings, and I'll briefly mention these.

AB Sciex - TOF-TOF

- **Models include: 4700 / 4800 / 5800**
- **The data files are stored in an Oracle database rather than as files**
- **Individual spectra can be exported as T2D file**
- **GPS Explorer no longer supported and has been replaced with Protein Pilot**
- **Can use Mascot or Paragon engine from PP.**

MASCOT : *Peak Picking*

© 2013 Matrix Science



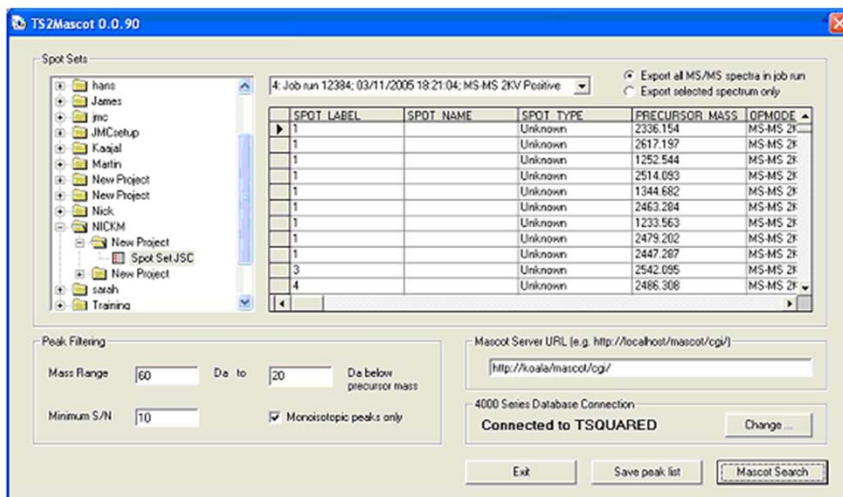
There are two data systems used by ABSciex instruments. One uses the Analyst software and generates wiff files, but the first type I will discuss is the TOF-TOF instrument which was first introduced back in 2001.

The data files are stored in an Oracle database rather than as files and although individual spectra can be exported as separate T2D files which can be opened by Distiller or processed in Mascot Daemon using the Data Explorer import filter, this tends to be inconvenient because of the huge number of files.

GPS Explorer was the original data processing software released with the instrument, but this is no longer supported by AB Sciex and has been replaced with Protein Pilot.

You can use Mascot or the Paragon engine from PP, which supports the TOF-TOF instrument and Wiff files. As you might expect, there is less functionality in PP when using Mascot compared with Paragon.

AB Sciex - TOF-TOF - TS2Mascot



MASCOT : Peak Picking

© 2013 Matrix Science



A free utility called TS2Mascot provides a convenient way to submit a search through the standard Mascot search form. This is a Windows program that communicates with the Oracle database using ODBC. The PC will need to have Oracle client software installed, including the Oracle ODBC driver. To submit a Mascot search from TS2Mascot, Microsoft Internet Explorer 4.01 SP2 or later must be installed and operational.

To install TS2Mascot, run the TS2Mascot installer. Unless 4000 Series Explorer or GPS Explorer is already installed on the target PC, you may need to create a local Oracle service name, as described in the setup instructions.

AB Sciex - TOF-TOF - MS Data Converter

- Support for Wiff and TOF-TOF data
- Free download from AB Sciex:
<http://www.absciex.com/MSDataConverter>
- The only way to use Distiller for SILAC type experiments with TOF-TOF data

MASCOT : *Peak Picking*

© 2013 Matrix Science

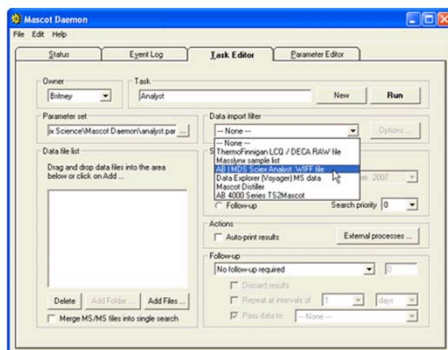


Another option with TOF-TOF data is to use the MS Data Converter which is a [free download](#) from AB Sciex. It supports both Wiff files and TOF/TOF data stored in an Oracle database. Output can be a simple MGF peak list or an mzML file containing a fairly complete representation of the raw data.

Converting a TOF/TOF spot set into an mzML file enables Mascot Distiller to be used for quantitation. Select the MS and MS/MS job runs corresponding to a single LC separation. If you have multiple separations, corresponding to fractions or replicates, each one must be exported to a separate mzML file, and you can then create a Distiller multi-file project for the set of files. The MS Data Converter User Guide explains how to format a path to specify the data to be exported. In most cases, you should choose to export profile data.

AB Sciex - wiff files - mascot.dll

- Supplied with recent versions of Analyst
- IE 7 or earlier when used from scripts menu



MASCOT : Peak Picking

© 2013 Matrix Science



For wiff files, a popular option is to use a script called “Mascot.dll”

This requires that a version of Analyst able to open the wiff files has been installed on the system. The script has been supplied with most versions of Analyst but it can also be downloaded from our website. If it’s been installed correctly, it will be available from the scripts menu in Analyst.

However, because of security restrictions in Internet Explorer, it requires IE 7 or earlier for auto filling in the file name and this will mean using Windows XP or earlier. To get around this restriction, it may be easier to use Mascot Daemon. So, if Mascot Daemon is installed on the same computer as Analyst, then Mascot.dll will be available as a download filter in the task editor tab in Mascot Daemon. A similar set of processing options will be available when using Mascot.dll from the scripts menu, or from Mascot Daemon.

Mascot.dll was written as a collaboration between Matrix Science and AB Sciex, but it is now supported and maintained by AB Sciex, so you may need to direct enquires about it to them.

AB Sciex - Triple TOF 5600

- Distiller is slow because of the huge data size, but no choice for quantitation
- MS Data Converter gives good results in a shorter amount of time, so for protein ID, can be a better choice.

MASCOT : *Peak Picking*

© 2013 Matrix Science



Unfortunately, the peak detection in Mascot Distiller is rather slow with 5600 data because of the huge number of spectra and the huge number of peaks per spectrum. You don't have any choice if you want to use Distiller for quantitation, but if you are just interested in protein ID, it may be worth considering using the MS Data Converter.

Agilent

- **Mass Hunter software supports export of MGF and mzData files**
- **Agilent files of all types are fully supported by Mascot Distiller**

MASCOT : *Peak Picking*

© 2013 Matrix Science



Recent Agilent instruments now all use the Mass Hunter software which supports the export of MGF and mzData files.

Many of our customers using Agilent instruments export individual mgf files from Mass Hunter and then combine these into a single search using Mascot Daemon.

We get good support and libraries from Agilent for opening and processing their data files, so all types of files from the instruments are fully supported by Mascot Distiller.

Bruker - BioTools / ProteinScape

- **BioTools and ProteinScape have full integration with Mascot Server**
- **Searches are easily submitted to a local server or public web site**
- **Search results automatically imported into BioTools / ProteinScape for further analysis**
- **Early versions required separate scripts on the Mascot server - no longer required.**

MASCOT : *Peak Picking*

© 2013 Matrix Science



BioTools has had full integration with Mascot Server since 1999, and Bruker work closely with us for each new Mascot Server release.

Searches are easily submitted to a local server or public web site. Search results automatically imported into BioTools / ProteinScape for further analysis. Early versions required separate scripts on the Mascot server – no longer required

There are different file formats for different Bruker instruments, and Mascot Distiller has support for reading all these files.

There is also a free command line tool, CompassXport that can be downloaded from the Bruker Web site.

Shimadzu / Kratos

- Support within Launchpad Software for submitting searches to Mascot
- Mascot Distiller opens and processes Axima and LCMSSolution files.

MASCOT : *Peak Picking*

© 2013 Matrix Science



There is support within the Shimadzu Launchpad Software for submitting searches to Mascot.

Mascot Distiller opens and processes Axima and LCMSSolution files.

Thermo

- File name is .RAW for all instruments
- Software is XCalibur
- Save survey scans as profile, ms-ms as centroided

- Large number of different options for producing the peak lists.

MASCOT : Peak Picking

© 2013 Matrix Science

 **MATRIX
SCIENCE**

For all Thermo instruments, the files saved are .RAW files, and this contains all the required information, so there is no need for any other file.

The Instrument control software is XCalibur, but this software cannot be used to submit searches to Mascot.

For LCMSMS experiments on the earlier LCQ instruments, it was common to generate a survey scan, an MS1 zoom scan for the region around the selected precursor, and then the MS2 scan – the so called “Triple Play” experiment. For all the more recent instruments, the survey scans have sufficient resolution for precursor mass and charge determination, so there is no need for zoom scans. If you have an older instrument and want to perform a quantitation experiment, for example SILAC, that uses MS1 scans, then you will need to perform a triple play experiment.

In general, the most efficient way to save the data is to save survey scans as profile data, and the ms-ms scans as ‘stick’ or ‘centroided’ data. With XCalibur 2.2 and the Q-Exactive instrument, there was originally no option to save the ms2 data in centroid mode. This was addressed in a version of the “Tune software” for Q Exactive (called Exactive series 2.2) which was released in April 2013.

Thermo - ExtractMSN

- Free software no longer supported by Thermo.
- Only for XCalibur versions prior v2.2
- Instructions on Matrix Science public web site for download and installation
http://www.matrixscience.com/help/instruments_xcalibur.html
- Also requires MSFileReader to be installed
- Previous names include lcq_dta.exe and extract_msn.exe

MASCOT : *Peak Picking*

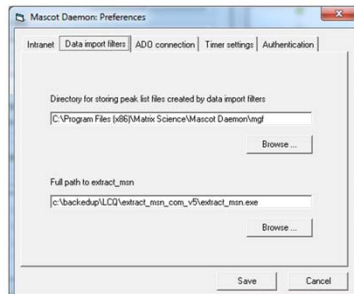
© 2013 Matrix Science



ExtractMSN, previously known as lcq_dta is free software that is unfortunately no longer supported by Thermo.

It will only work for XCalibur versions prior to version 2.2. There are instructions on our public web site for download and installation and please note that you will also need to follow the instructions to install MSFileReader

Thermo - ExtractMSN - From Daemon



If you get

'No output from extract_msn.exe (check parameters)'

Run from the command line:

```
"C:\Program Files\Thermo\ExtractMSN\extract_msn.exe" -G1 TEST.RAW
```

MASCOT : Peak Picking

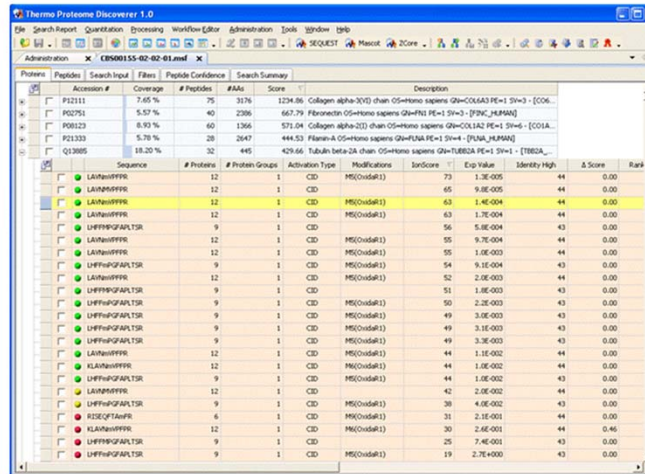
© 2013 Matrix Science



Once you have got it ExtractMSN.exe installed, you can set up Daemon to use it. For older versions of Daemon, you have to rename extract_msn.exe on the disk.

A common problem is getting this error message: 'No output from extract_msn.exe (check parameters)' and the best way to resolve this is to try and run extract_msn from the command line. Open a command prompt, change to the directory that has the raw file using the CD command and then run the command as shown here, replacing TEST.RAW with the name of your file. You should see garbage on the screen, possibly hear beeps and when it's finished, there should be a stack of .dta files in the current directory.

Thermo - Proteome Discoverer



The screenshot displays the Thermo Proteome Discoverer 1.0 interface. The top menu bar includes options like Search Report, Quantitation, Processing, Workflow Editor, Administration, Tools, Window, and Help. The main window shows a search summary for a file named 'CE060155-02-02-01.msf'. Below this, there are two tables. The first table lists search results with columns for Accession #, Coverage, # Peptides, #MS, and Score. The second table is a detailed list of peptides with columns for Sequence, # Proteins, # Protein Groups, Activation Type, Modifications, IonScore, Exp Value, Identity High, Δ Score, and Rank. The peptide 'LAIHNSPPFR' is highlighted in yellow, showing a score of 65 and an identity high of 44.

Accession #	Coverage	# Peptides	#MS	Score	Description
P12111	7.65 %	75	1176	1234.86	Collagen alpha-2(XI) chain OS=Homo sapiens GI=55643 PE=1 SW=3 - [COLL_
P02751	5.57 %	40	2386	667.79	Fibrinectin OS=Homo sapiens GI=191 PE=1 SW=3 - [FN1_HUMAN]
P08123	8.93 %	60	1366	571.04	Collagen alpha-2(I) chain OS=Homo sapiens GI=55643 PE=1 SW=3 - [COL1A_
P12133	5.78 %	28	2847	444.53	Fibrin alpha-2 chain OS=Homo sapiens GI=191 PE=1 SW=3 - [FIB1A_HUMAN]
Q13885	18.20 %	32	445	429.66	Tubulin beta-2A chain OS=Homo sapiens GI=1472 PE=1 SW=1 - [TUBB2A_

Sequence	# Proteins	# Protein Groups	Activation Type	Modifications	IonScore	Exp Value	Identity High	Δ Score	Rank
LAIHNSPPFR	12	1	CID	MS(Oxidat1)	73	1.3E-005	44	0.00	
LAIHNSPPFR	12	1	CID	MS(Oxidat1)	65	9.8E-005	44	0.00	
LAIHNSPPFR	12	1	CID	MS(Oxidat1)	63	2.4E-004	44	0.00	
LAIHNSPPFR	12	1	CID	MS(Oxidat1)	63	1.7E-004	44	0.00	
LHFFHSQAFALTSR	9	1	CID	MS(Oxidat1)	56	5.8E-004	43	0.00	
LAIHNSPPFR	12	1	CID	MS(Oxidat1)	55	9.7E-004	44	0.00	
LAIHNSPPFR	12	1	CID	MS(Oxidat1)	55	1.0E-003	44	0.00	
LHFFHSQAFALTSR	9	1	CID	MS(Oxidat1)	54	9.1E-004	43	0.00	
LAIHNSPPFR	12	1	CID	MS(Oxidat1)	52	2.0E-003	44	0.00	
LHFFHSQAFALTSR	9	1	CID	MS(Oxidat1)	51	1.8E-003	43	0.00	
LHFFHSQAFALTSR	9	1	CID	MS(Oxidat1)	50	2.0E-003	43	0.00	
LHFFHSQAFALTSR	9	1	CID	MS(Oxidat1)	49	3.0E-003	43	0.00	
LHFFHSQAFALTSR	9	1	CID	MS(Oxidat1)	49	3.1E-003	43	0.00	
LHFFHSQAFALTSR	9	1	CID	MS(Oxidat1)	49	3.3E-003	43	0.00	
LAIHNSPPFR	12	1	CID	MS(Oxidat1)	44	1.1E-002	44	0.00	
KLAHNSPPFR	12	1	CID	MS(Oxidat1)	44	1.0E-002	44	0.00	
LHFFHSQAFALTSR	9	1	CID	MS(Oxidat1)	44	1.0E-002	43	0.00	
LAIHNSPPFR	12	1	CID	MS(Oxidat1)	42	2.0E-002	44	0.00	
LHFFHSQAFALTSR	9	1	CID	MS(Oxidat1)	38	4.0E-002	43	0.00	
RSEIQAVFR	6	1	CID	MS(Oxidat1)	31	2.1E-001	44	0.00	
KLAHNSPPFR	12	1	CID	MS(Oxidat1)	30	2.6E-001	44	0.46	
LHFFHSQAFALTSR	9	1	CID	MS(Oxidat1)	25	7.4E-001	43	0.00	
LHFFHSQAFALTSR	9	1	CID	MS(Oxidat1)	19	2.7E+000	43	0.00	

MASCOT : Peak Picking

© 2013 Matrix Science



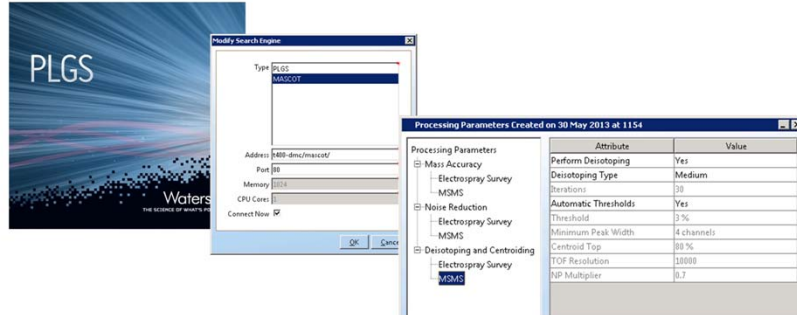
Thermo's Proteome Discoverer provides fully automated raw file processing and search submission. Peak picking and search parameters are selected in a workflow wizard. When the search is complete, the results are imported into Proteome Discoverer, where they can be filtered and inspected.

(Note that the local Mascot Server URL must be entered in the form <http://ec-vm2/mascot/> where ec-vm2 is replaced by the hostname of your local server.)

You can use the demo version of Proteome Discoverer to generate MGF files. The download and the description can be found on the Proteomics Portal (<http://portal.thermo-brims.com>) – you need to create an account to get the download, and you need to set up any workflows **before** the demo expires.

There are a number of other options for Thermo data, including DeconMS, DTASuperCharge, a Merge.pl script that we provide and RAW2MSM. These are all detailed on our public web site.

Waters



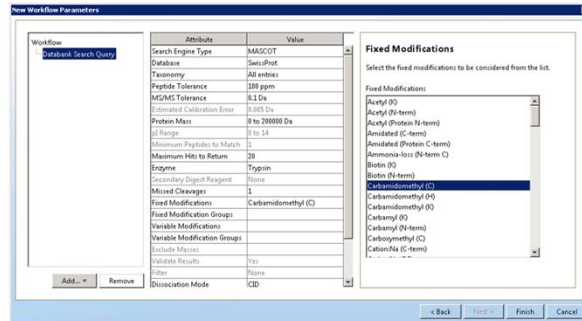
MASCOT : Peak Picking

© 2013 Matrix Science



For Waters data, Protein Lynx Global Server can be used for peak detection and submitting searches to either PLGS or Mascot. Simply give the URL for your Mascot Server, (without the http:// prefix). Choose suitable peak detection parameters as part of your workflow

Waters



- Alternatively, Distiller is an option, but this currently does not support MS^E data

MASCOT : Peak Picking

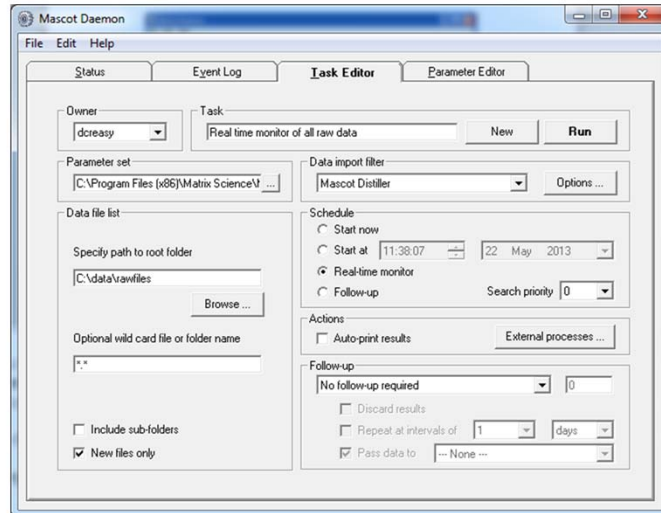
© 2013 Matrix Science



And then choose the Mascot Search parameters.

Alternatively, Distiller is an option, but this currently does not support MS^E data

Mascot Daemon



MASCOT : Peak Picking

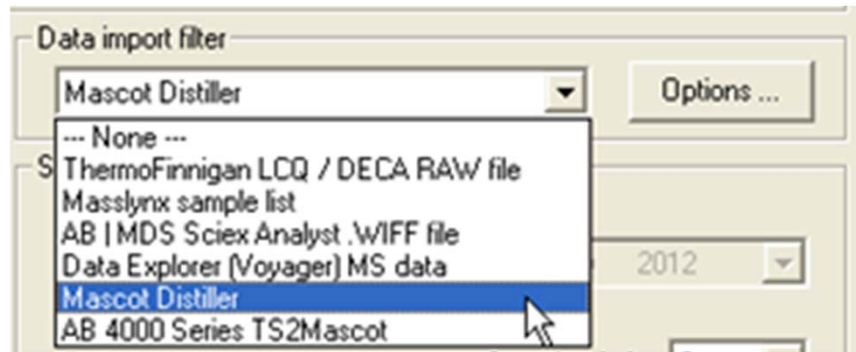
© 2013 Matrix Science



So, that's the last of the instrument vendor options.

Mascot Daemon is supplied at no additional cost with Mascot Server and is our tool for automating peak picking and searching data. It takes raw files, or text files such as mgf, or mzML files and submits them to your in-house Mascot server using predefined parameters. Although the first protein match from a search is returned in Mascot Daemon, it isn't a tool for reviewing results, and there is a link to the results from the status tab to enable you to view these in your web browser.

Mascot Daemon - Import filters



MASCOT : Peak Picking

© 2013 Matrix Science

MATRIX
SCIENCE

If you don't specify an import filter, then Mascot Daemon will be expecting text files, for example mgf files, as specified in the Daemon parameter editor. This can be useful if your instrument software is able to export mgf or mzML files automatically. If you are using mgf files, there is an option to get Mascot Daemon to merge all these files and submit a single large search.

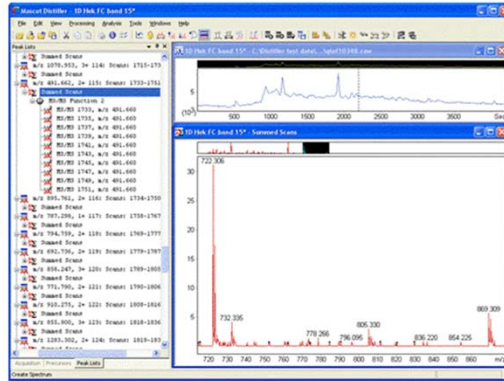
The other options include

- Extract-msn, which is, for historical reasons, ThermoFinnigan LCQ/DECA RAW file in the list.
- Masslynx sample list is no longer useful for most people.
- Data Explorer (Voyager) MS Data option works for both the old Voyager instruments, and for the T2D files from the TOF-TOF as described earlier.

I'll discuss Distiller next, but finally, we have the TS2Mascot option that I've already described.

Distiller - advantages and disadvantages

- + Consistent interface for all instruments
- + Automated using Mascot Daemon
- + High quality peak lists
- Not free
- Can be slow



MASCOT : Peak Picking

© 2013 Matrix Science

MATRIX
SCIENCE

There are advantages and disadvantages with using Distiller and it may not be the right solution for everyone. For example, with some instruments, it can be quite slow, but even with those instruments, the higher quality peak lists or consistent workflow may justify the extra time and cost required.

For labs with instruments from different manufacturers, there is the advantage of having the same user interface, same peak detection protocols and same quantitation methods for all instruments. Also, the process of peak detection, search submission and quantitation can be automated using Mascot Daemon. Mascot Distiller produces high quality peak lists, but there is a cost involved, and the process can be slow

Distiller - peak list creation

Options to:

- re-determine the precursor mass
- re-determine the precursor charge
- 'de-charge' fragments and precursors
- handle reporter ions for iTRAQ, TMT etc.
- output in multiple formats.

Apart from creating high quality peak lists, that is peak lists with fewer noise peaks, the following things will help get better Mascot search results:

The precursor mass and charge in the raw data file need to be determined very quickly by the instrument control software, and these will sometimes be incorrect. In Distiller, these can be re-determined from the survey scans.

Mascot will only look for fragments with a charge of 1 or 2 which is sufficient for many instruments. However, there's an option to 'de-charge' higher charge state fragments where the data have high enough resolution.

Of, course, Distiller is able to pass through iTRAQ or TMT peaks so that quantitation data is correct.

And finally, there is support for multiple peak list formats including mgf, dtaand mzData.

Distiller - choosing parameters

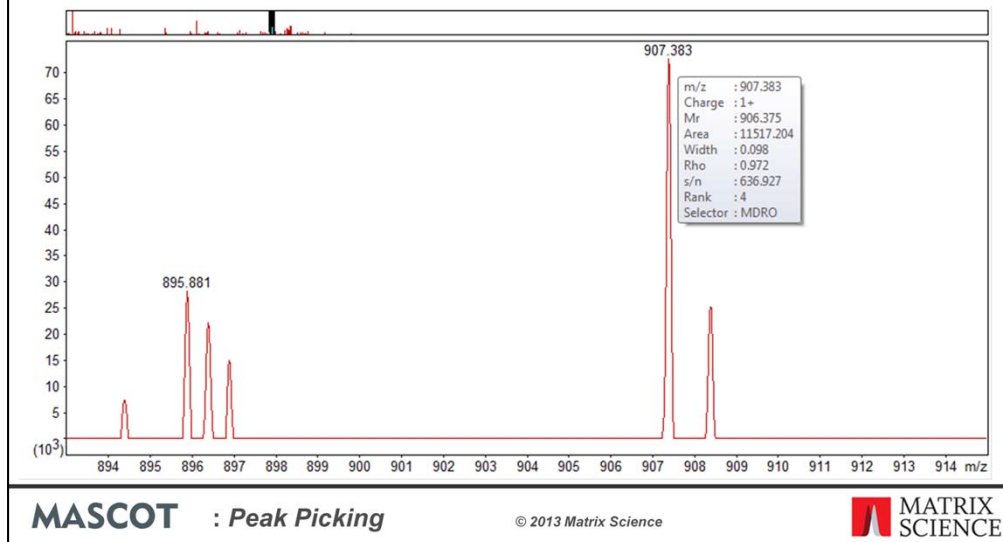
- We supply 42 options files
- For a new instrument, or a novel method, send us a file and we can optimise params
- Look at peak detection results in Distiller:

We currently supply 42 different options files. You don't see all of these in the dropdown list when you load the raw data file because they are filtered by instrument manufacturer or type.

If you have recently released instrument and ask for help, we may ask you to send a file for a standard so we can propose suitable parameters.

The most important thing however, is to look at the peak detection results in Distiller.

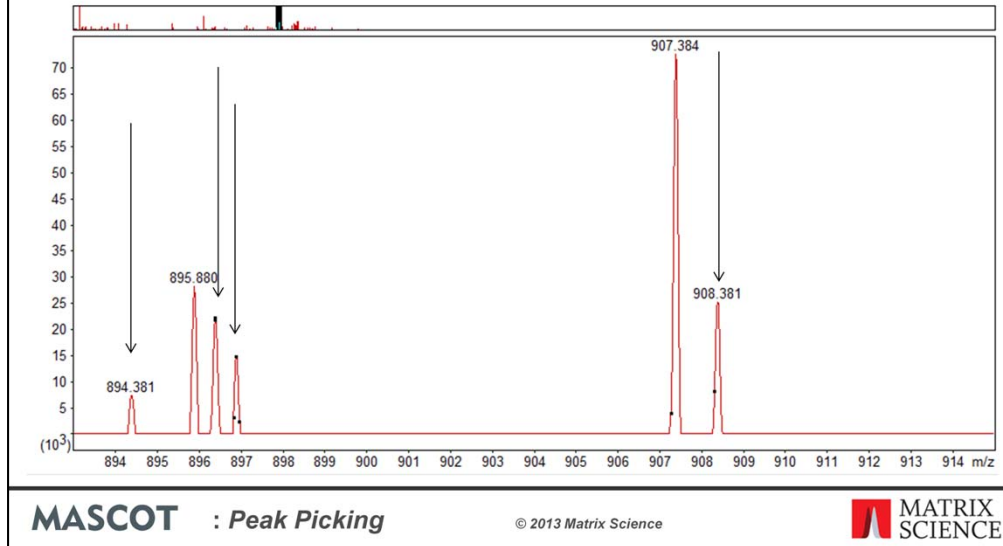
Checking peak detection parameters



These tips don't just apply to Distiller. Whatever software you are using, check out a few spectra, and try zooming in on some peaks.

For example, here in Distiller, we see all is working well. This first, lone peak is clearly not an isotopic cluster, and has not been labelled as a peak. Just the ¹²C peak of the next two peaks have been detected, and in both cases the charge is detected correctly.

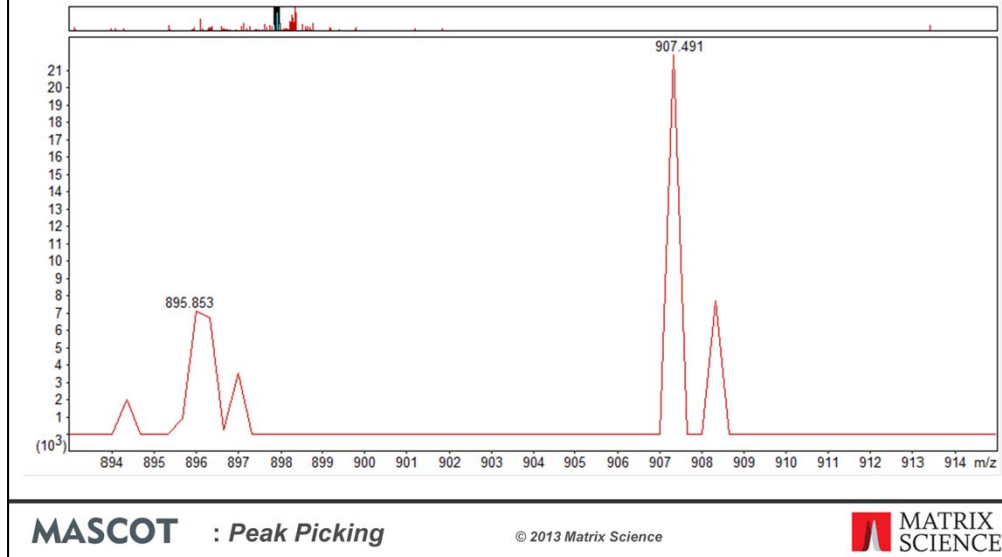
Checking peak detection parameters



But if we got the parameters very wrong, we'd maybe see an extra peak labelled here, and the ^{13}C peaks all labelled.

If there's no room for a label, distiller just puts a small black square to indicate a peak.

Checking peak detection parameters



Apart from peak detection, there are other parameters that affect how the data is shown and prepared before peak detection. If, instead of the lovely peaks you saw in the previous slide, you see these saw tooth shaped peaks, then you will know something is wrong with the regridding parameters in Distiller.

In this case, it would make sense to compare a couple of spectra in XCalibur with those in Distiller before changing parameters, or asking for support from us.

There is extensive help in Distiller for peak detection parameters, but if you have problems, please don't hesitate to ask for help.

Checking peak detection parameters

Options file	Processing time (h:mm:ss)	Matches with score>100	Significant at 1% FDR	Significant and 13C	Queries	Peak list size (MB)
Orbitrap_low_res_MS2_1.opt	3:32:33	258	2958	1.6%	9791	40.8
Orbitrap_low_res_MS2_2.opt	2:59:31	258	2959	1.6%	9785	40.8
Orbitrap_low_res_MS2_3.opt	2:52:30	262	2970	49.4%	9990	41.4
Orbitrap_low_res_MS2_4.opt	0:12:22	118	2839	0.9%	9786	73.3
extract_msn.opt	0:00:40	117	2886	49.4%	9990	74.2

MASCOT : Peak Picking

© 2013 Matrix Science



Even within Distiller, there is a trade off between processing time and quality of results. There is an example in the help for different options for XCalibur files. In this case however, you would only get an extra 2.5% matches at 1%FDR if you are prepared to wait 3 and half hours instead of 40 seconds for peak processing – probably not worth it!

But, the exercise is probably worth repeating with your own data. Spending some time investigating parameters up front can pay good dividends.

Distiller - automation using Daemon

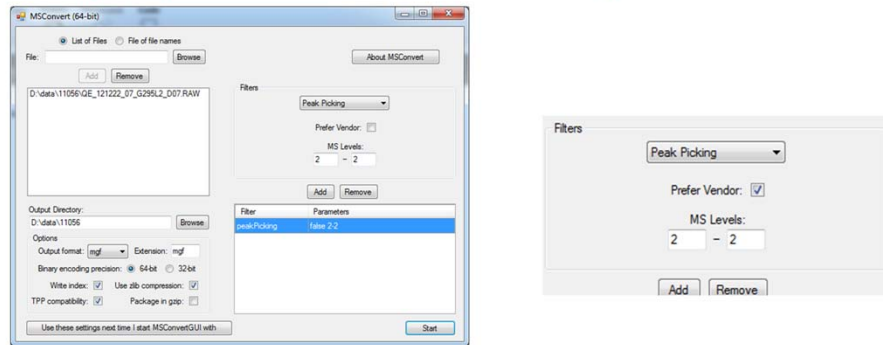
- Requires MD-GUI and MD-DTB
- Multiple threads can be enabled - see help: *Multi-file Quantitation*
- Parameter files

To be able to use Distiller peak picking from Mascot Daemon requires that, as a minimum, you have Mascot Distiller workstation and the Daemon toolbox on the computer with Mascot Daemon.

By default, when Daemon uses Distiller for peak picking, it will only use a single thread for each file. This is the same behaviour as in Distiller 2.3 and earlier. The reason is that Daemon is a 32-bit application, and the more threads that are used, the more likely that Daemon will run out of address space when processing very large files or processing multiple files in parallel in separate tasks. If you wish to use more threads for peak picking on an experimental basis, this behaviour can be changed using a registry key. See the Mascot Distiller help, *Multi-file Quantitation*

Third party software tools - MSConvert

<http://proteowizard.sourceforge.net/>



MASCOT : Peak Picking

© 2013 Matrix Science



The most widely used free software for converting raw files to peak lists is probably now Proteowizard, which includes a utility called MSConvert.

There used to be separate tools from TPP, the trans proteomic pipeline that comes from the Institute for Systems Biology in Seattle, but they now also use MSConvert. There is a simple GUI interface to this, as shown here. But, as their help says, the GUI is not always kept up to date with the latest filter options, so it is better to use the command line instead.

There is limited documentation for the peak picking in MSConvert, and the recommendation is to use the 'Vendor library' option. This means that, for example, with a Thermo file, it will use Thermo's peak picking code provided in a DLL supplied by Thermo. Some of the vendors don't provide this code, but Thermo does. So, for a file where the ms-ms data has been saved as profile data, it's recommended that you use 'Vendor' peak picking.

Conclusions

- Many choices, but worth spending time comparing results before deciding
- Consider quantitation needs
- Refine one protocol, and then stick to it
- Additional help on our web site.