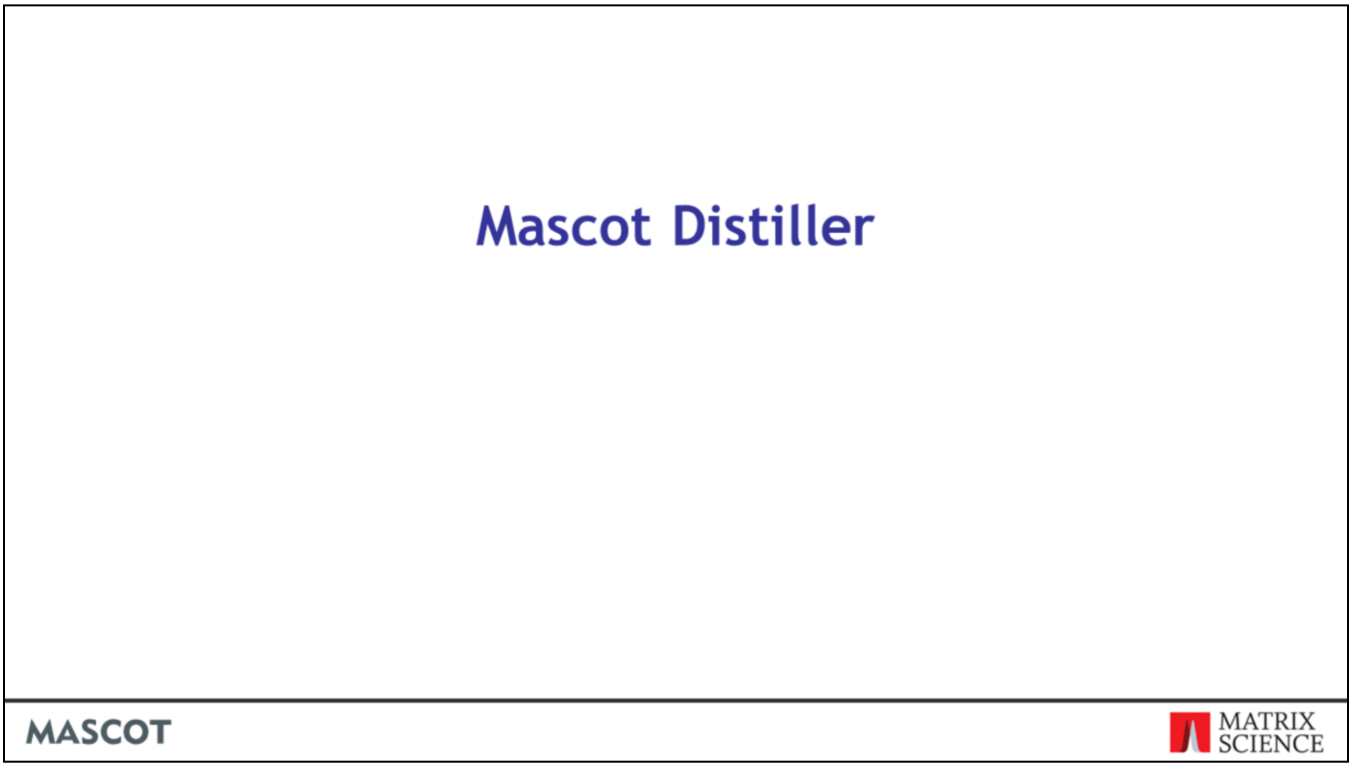
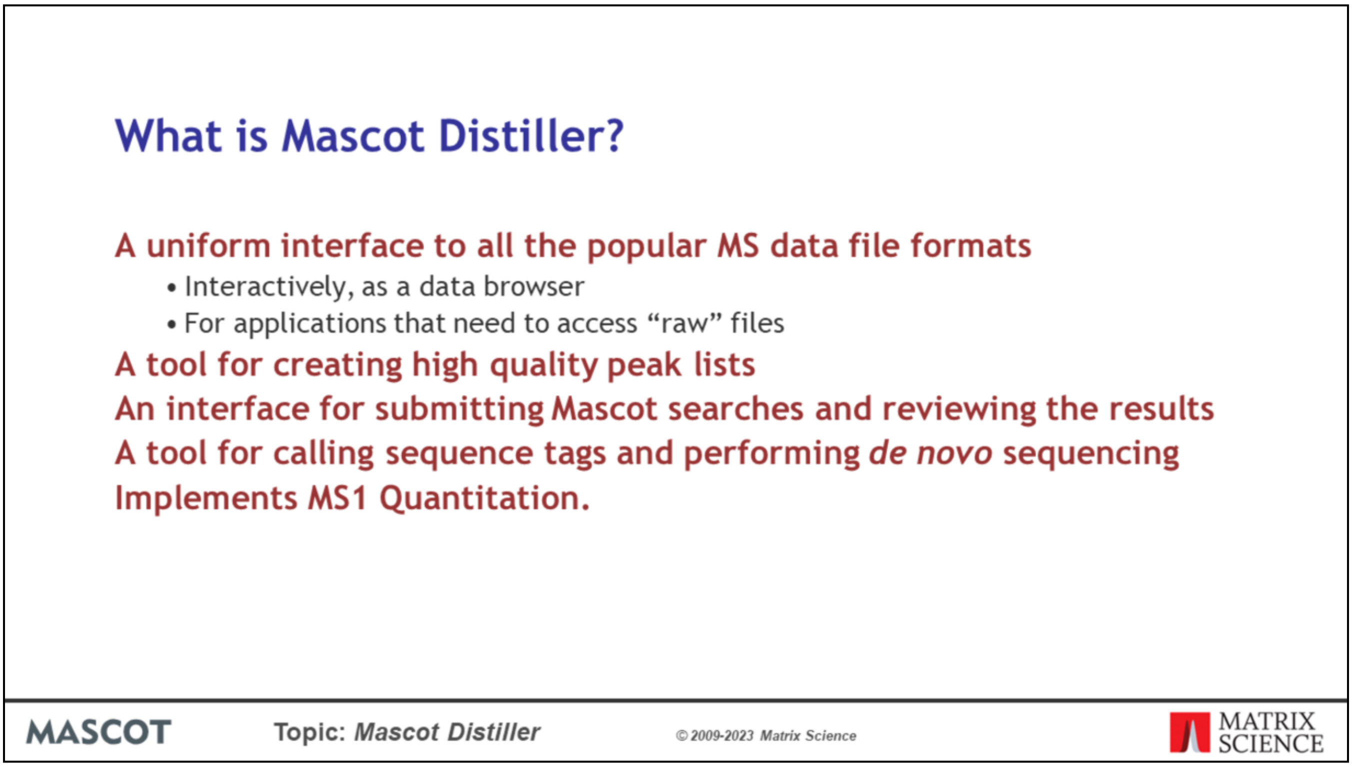


Mascot Distiller



What is Mascot Distiller?

A uniform interface to all the popular MS data file formats

- Interactively, as a data browser
- For applications that need to access “raw” files

A tool for creating high quality peak lists

An interface for submitting Mascot searches and reviewing the results

A tool for calling sequence tags and performing *de novo* sequencing

Implements MS1 Quantitation.

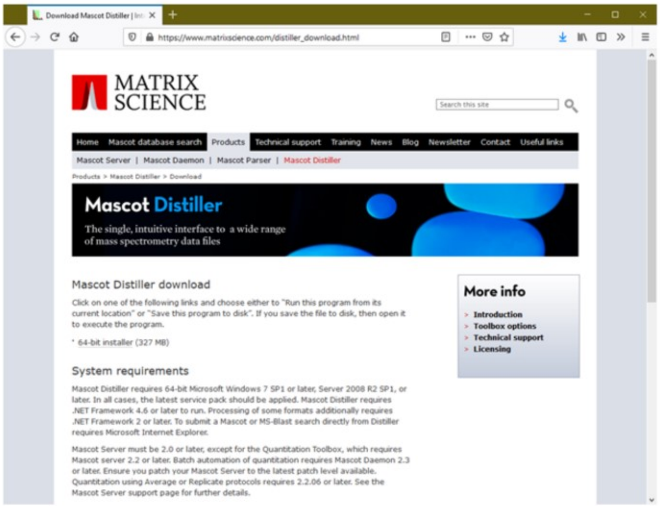
Most laboratories will have instruments from more than one manufacturer. The instrument data systems are necessarily complex, so there can be a steep learning curve for someone who comes into the lab and just wants to browse their data or generate peak lists. The first benefit of Mascot Distiller is that you can access all of the popular data formats from a single user interface.

Another reason for developing Distiller was to produce high quality peak lists without having to constantly tweak peak detection parameters. Poor quality peak lists translate into poor quality Mascot scores.

Distiller is also a powerful way to review Mascot search results. And, if Mascot fails to get a match, you can perform *de novo* sequencing and interpret sequence tags for tag searches.

Finally, Distiller is used for quantitation methods that require information from the raw data file, either because it is necessary to integrate the elution profile of each precursor peptide or because information is required for precursor peptides that were not used to trigger MS/MS scans, so are missing from the peak list. There is a separate presentation dealing with quantitation.

What is Mascot Distiller?


Download Mascot Distiller | Int
https://www.matrixscience.com/distiller_download.html
MATRIX SCIENCE
Search this site
Home
Mascot database search
Products
Technical support
Training
News
Blog
Newsletter
Contact
Useful links
Mascot Server | Mascot Daemon | Mascot Parser | Mascot Distiller
Products > Mascot Distiller > Download
Mascot Distiller
The single, intuitive interface to a wide range of mass spectrometry data files
More info
Introduction
Toolbox options
Technical support
Licensing

Mascot Distiller download

Click on one of the following links and choose either to "Run this program from its current location" or "Save this program to disk". If you save the file to disk, then open it to execute the program.

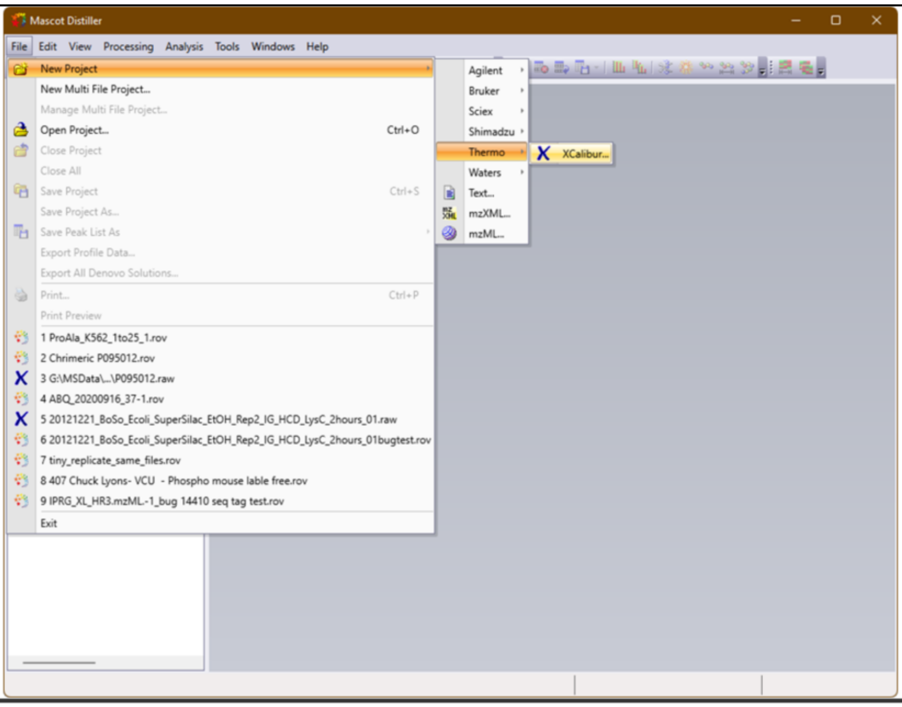
- 64-bit installer (327 MB)

System requirements

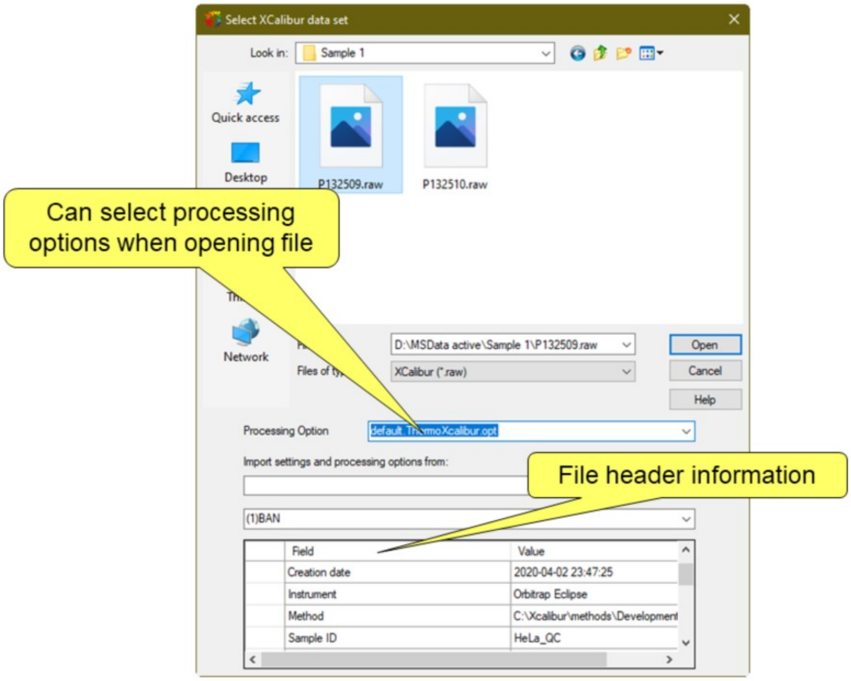
Mascot Distiller requires 64-bit Microsoft Windows 7 SP1 or later, Server 2008 R2 SP1, or later. In all cases, the latest service pack should be applied. Mascot Distiller requires .NET Framework 4.6 or later to run. Processing of some formats additionally requires .NET Framework 2 or later. To submit a Mascot or MS-Blast search directly from Distiller requires Microsoft Internet Explorer.

Mascot Server must be 2.0 or later, except for the Quantitation Toolbox, which requires Mascot server 2.2 or later. Batch automation of quantitation requires Mascot Daemon 2.3 or later. Ensure you patch your Mascot Server to the latest patch level available. Quantitation using Average or Replicate protocols requires 2.2.06 or later. See the Mascot Server support page for further details.

If you are not already a Distiller user, you can get a free 30 day evaluation licence. Details are on our web site, under Products; Mascot Distiller.


Mascot Distiller
File
Edit
View
Processing
Analysis
Tools
Windows
Help
New Project
New Multi File Project...
Manage Multi File Project...
Open Project... Ctrl+O
Close Project
Close All
Save Project Ctrl+S
Save Project As...
Save Peak List As
Export Profile Data...
Export All Denovo Solutions...
Print... Ctrl+P
Print Preview
1 ProAla_K562_1to25_1.rov
2 Chrimeric P095012.rov
3 G:\MSData\...\P095012.raw
4 ABQ_20200916_37-1.rov
5 20121221_BoSo_Ecoli_SuperSilac_EtOH_Rep2_IG_HCD_LysC_2hours_01.raw
6 20121221_BoSo_Ecoli_SuperSilac_EtOH_Rep2_IG_HCD_LysC_2hours_01bugtest.rov
7 tiny_replicate_same_files.rov
8 407 Chuck Lyons- VCU - Phospho mouse lable free.rov
9 IPRG_XL_HR3.mzML-1_bug 14410 seq tag test.rov
Exit
Agilent
Bruker
Sciex
Shimadzu
Thermo
Waters
Text...
mzXML...
mzML...
XCalibur...

A binary or "raw" file is initially opened as a new Distiller project. We'll run through the supported formats in a minute.



Look in: Sample 1

P132509.raw

P132510.raw

File name: D:\MSData active\Sample 1\P132509.raw

Files of type: XCalibur (*.raw)

Processing Option: default_ThermoXcalibur.opt

Import settings and processing options from:

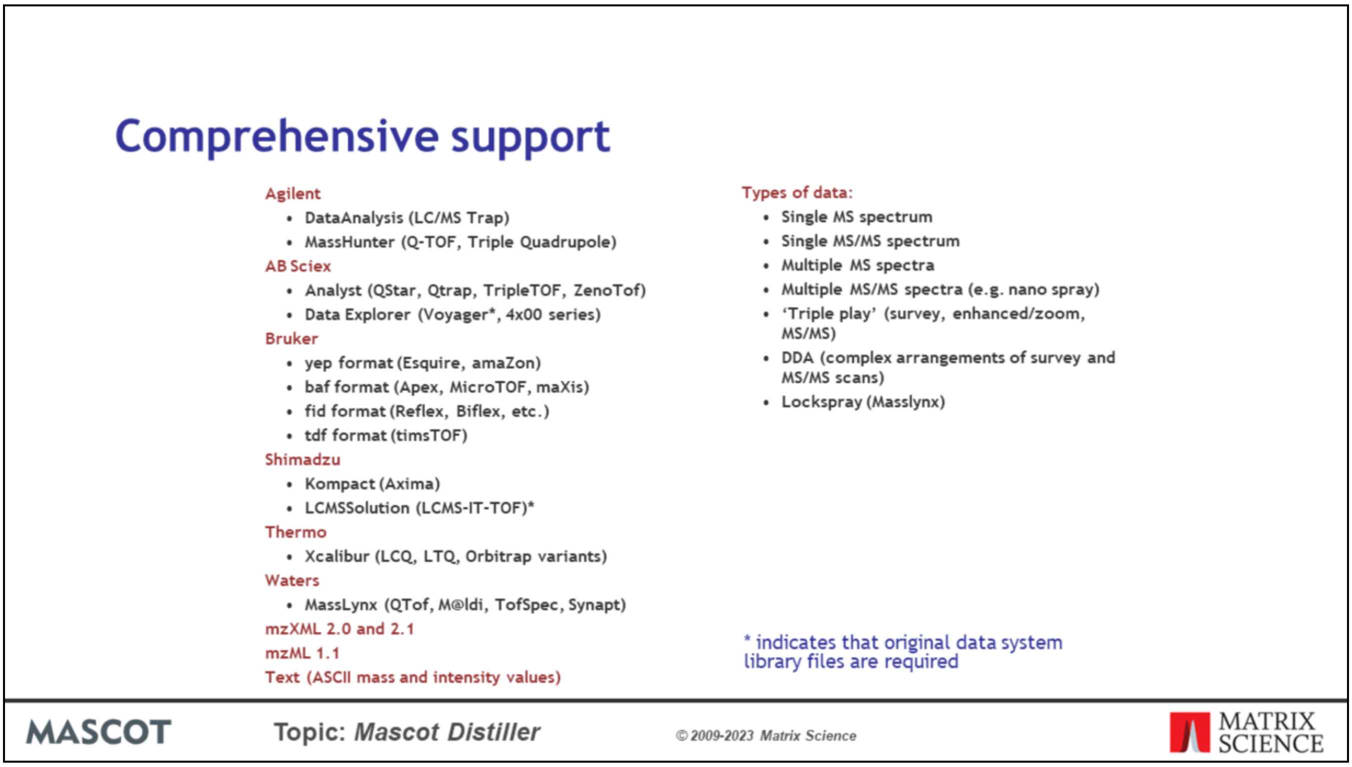
(1)BAN

Field	Value
Creation date	2020-04-02 23:47:25
Instrument	Orbitrap Eclipse
Method	C:\Xcalibur\methods\Development
Sample ID	HeLa_QC

Can select processing options when opening file

File header information

The file browse box displays some header information for the selected file. Also, this is where you can choose the processing options. You can change these later, but choosing the correct set here saves a couple of mouse clicks.



Comprehensive support

Agilent

- DataAnalysis (LC/MS Trap)
- MassHunter (Q-TOF, Triple Quadrupole)

AB Sciex

- Analyst (QStar, Qtrap, TripleTOF, ZenoTof)
- Data Explorer (Voyager*, 4x00 series)

Bruker

- yep format (Esquire, amaZon)
- baf format (Apex, MicroTOF, maXis)
- fid format (Reflex, Biflex, etc.)
- tdf format (timsTOF)

Shimadzu

- Kompact (Axima)
- LCMSSolution (LCMS-IT-TOF)*

Thermo

- Xcalibur (LCQ, LTQ, Orbitrap variants)

Waters

- MassLynx (QTof, M@ldi, TofSpec, Synapt)

mzXML 2.0 and 2.1

mzML 1.1

Text (ASCII mass and intensity values)

Types of data:

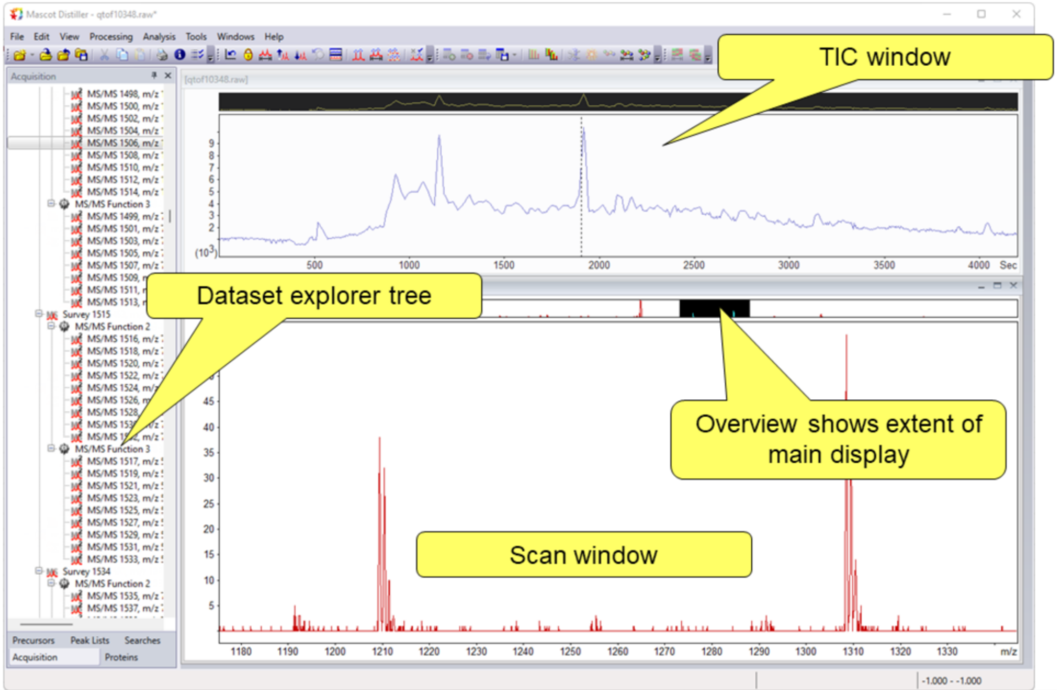
- Single MS spectrum
- Single MS/MS spectrum
- Multiple MS spectra
- Multiple MS/MS spectra (e.g. nano spray)
- 'Triple play' (survey, enhanced/zoom, MS/MS)
- DDA (complex arrangements of survey and MS/MS scans)
- Lockspray (Masslynx)

* indicates that original data system library files are required

Mascot Distiller supports all of the mainstream data file formats. In a few cases, Distiller requires library files that are installed as part of the instrument operating system.

Data files can be as simple as a single MS or MS/MS scan, or they can be the most complex mixtures of dependent scans created by IDE-type experiments.

There are some types of data that Mascot Distiller cannot handle. For example, it wouldn't know what to do with MRM (multiple reaction monitoring) data.


Mascot Distiller - qtof10348.raw*
File Edit View Processing Analysis Tools Windows Help
Acquisition
TIC window
Dataset explorer tree
Overview shows extent of main display
Scan window
Survey 1515
MS/MS Function 2
MS/MS Function 3
Survey 1534
Precursors Peak Lists Searches
Acquisition Proteins
500 1000 1500 2000 2500 3000 3500 4000 Sec
1180 1190 1200 1210 1220 1230 1240 1250 1260 1270 1280 1290 1300 1310 1320 1330 m/z

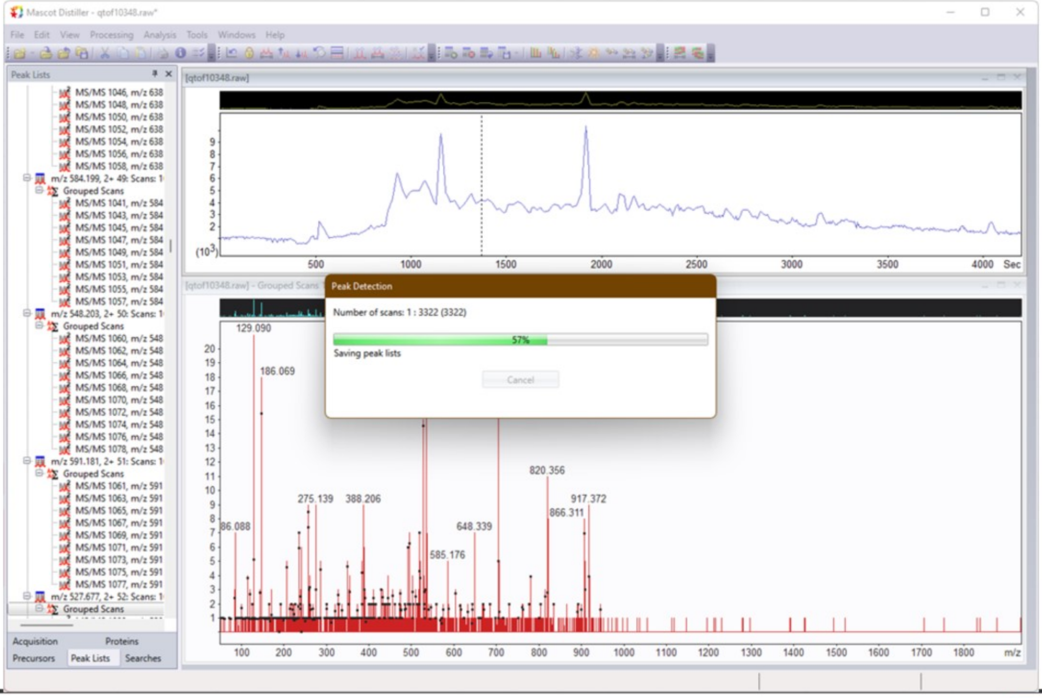
If the raw file contains LC-MS/MS data, this is the general appearance of the Distiller screen.

The Acquisition tab on the explorer tree shows the scan structure. This is Masslynx, so each survey scan is followed by MS/MS scans grouped into functions. If we were looking at (say) Xcalibur triple play data or Analyst data, this structure would look very different.

The total ion chromatogram window is chiefly a navigational aid.

The scan window displays a mass spectrum trace, selected by clicking on the explorer tree or the TIC trace.

At the top of the TIC and scan windows, you can see a representation of the whole trace, called the overview. The black area shows the portion of the trace that is currently displayed. This can be dragged or resized to make zooming and panning around the trace very easy.



Peak Detection

Number of scans: 1 : 3322 (3322)

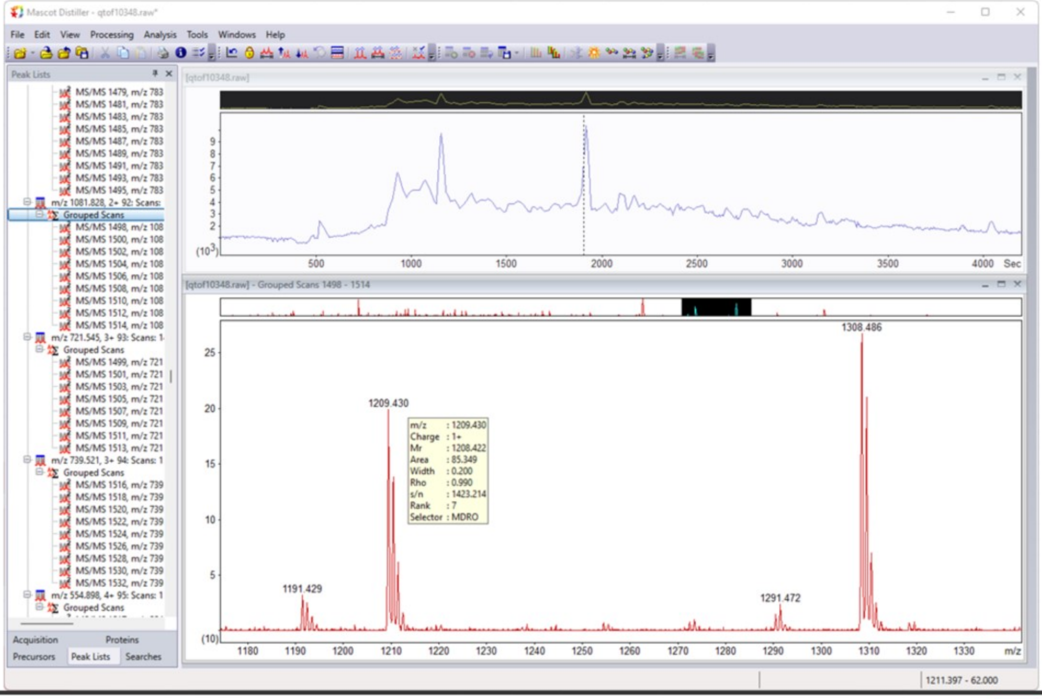
57%

Saving peak lists

Cancel

MASCOT **Topic:** ***Mascot Distiller*** *© 2009-2023 Matrix Science* MATRIX SCIENCE

To process the raw data into peak lists suitable for database searching, we just choose process from the Processing menu or toolbar. We can process the current scan, the currently displayed scan range, or all scans.



MS/MS 1479, m/z 783
MS/MS 1481, m/z 783
MS/MS 1483, m/z 783
MS/MS 1485, m/z 783
MS/MS 1487, m/z 783
MS/MS 1489, m/z 783
MS/MS 1491, m/z 783
MS/MS 1493, m/z 783
MS/MS 1495, m/z 783
m/z 1081.828, 2+ 92: Scans:
Grouped Scans
MS/MS 1498, m/z 108
MS/MS 1500, m/z 108
MS/MS 1502, m/z 108
MS/MS 1504, m/z 108
MS/MS 1506, m/z 108
MS/MS 1508, m/z 108
MS/MS 1510, m/z 108
MS/MS 1512, m/z 108
MS/MS 1514, m/z 108
m/z 721.545, 3+ 93: Scans: 1
Grouped Scans
MS/MS 1499, m/z 721
MS/MS 1501, m/z 721
MS/MS 1503, m/z 721
MS/MS 1505, m/z 721
MS/MS 1507, m/z 721
MS/MS 1509, m/z 721
MS/MS 1511, m/z 721
MS/MS 1513, m/z 721
m/z 739.521, 3+ 94: Scans: 1
Grouped Scans
MS/MS 1516, m/z 739
MS/MS 1518, m/z 739
MS/MS 1520, m/z 739
MS/MS 1522, m/z 739
MS/MS 1524, m/z 739
MS/MS 1526, m/z 739
MS/MS 1528, m/z 739
MS/MS 1530, m/z 739
MS/MS 1532, m/z 739
m/z 554.898, 4+ 95: Scans: 1
Grouped Scans

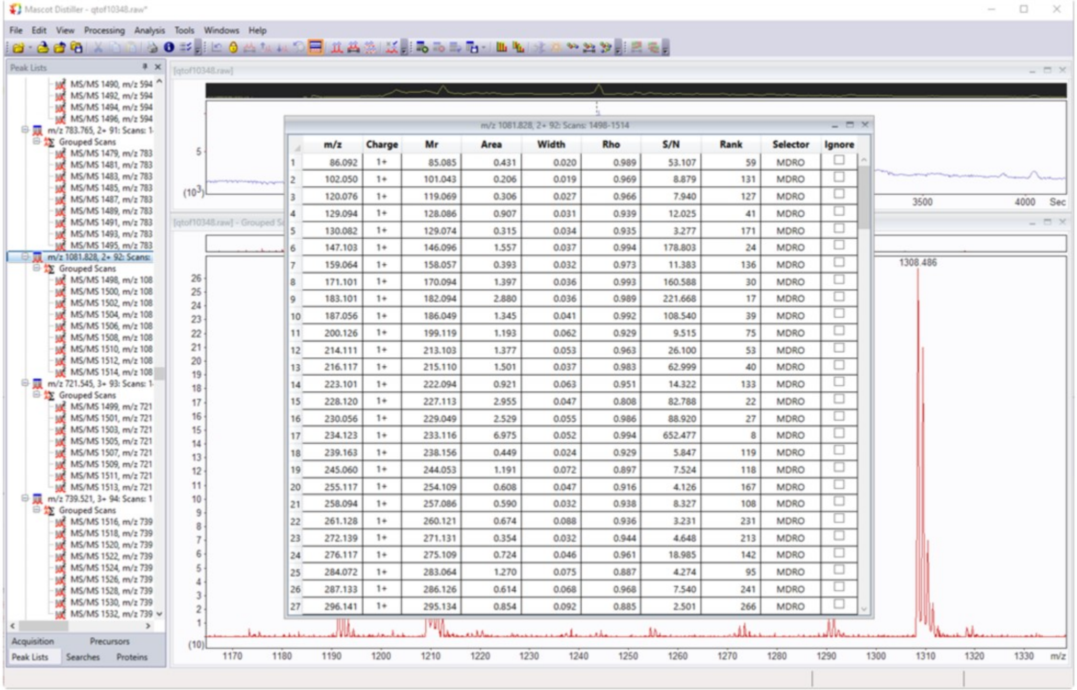
m/z	: 1209.430
Charge	: 1+
Mr	: 1208.422
Area	: 85.349
Width	: 0.200
Rho	: 0.990
s/n	: 1423.214
Rank	: 7
Selector	: MDRO

After processing, there is a new tab on the explorer tree: Peak lists. This shows the new data structure, which will usually be different from the original acquisition structure, because scans from the same precursor have been summed together.

When the mouse cursor is over a peak label, we get a tooltip showing complete information about the peak. m/z, charge, Mr, and area are fairly obvious. Width is the full width at half maximum height in Daltons.

Rho is the correlation coefficient, which measures the quality of the peak. Anything over 0.7 is normally a real peak and not a noise spike.

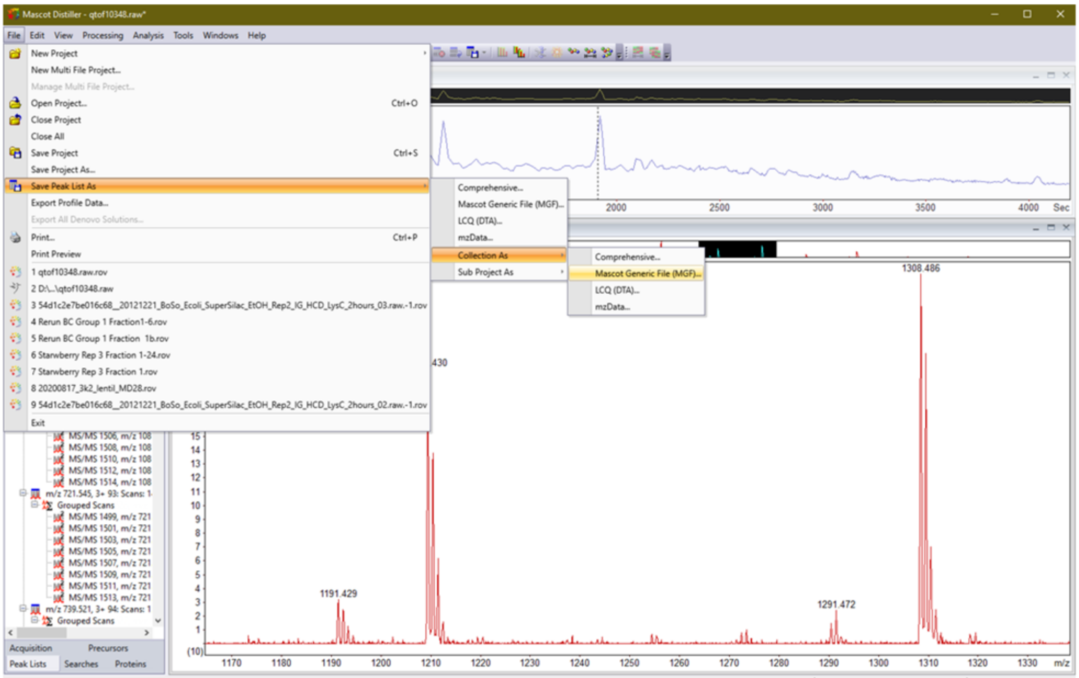
The peak picking procedure also gives us a signal to noise ratio for each peak. Rank is just the order in which the peak was picked and selector shows that this peak was picked by the Distiller library, which we call MDRO (from Mascot Data Reduction Object), as opposed to a manually edited peak



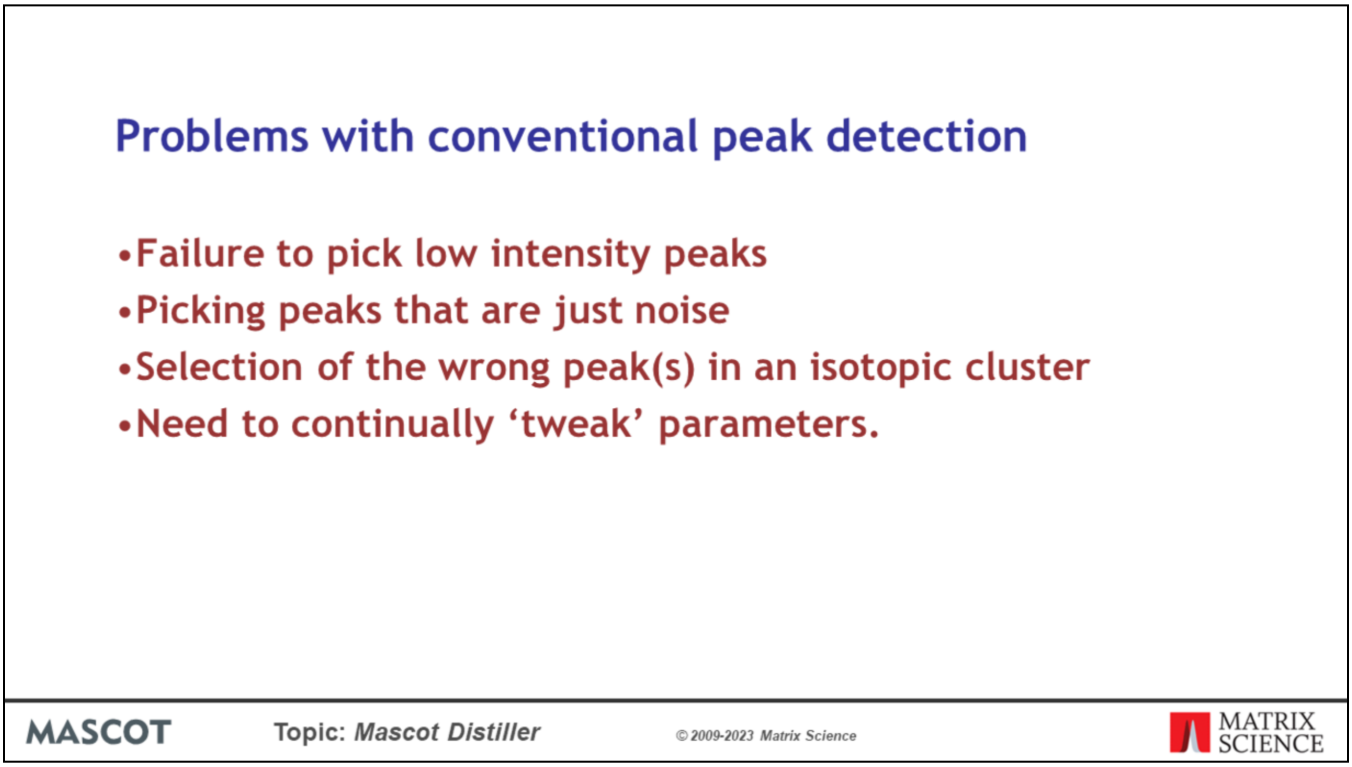
m/z 1081.828, 2+ 92: Scans: 1498-1514

	m/z	Charge	Mr	Area	Width	Rho	S/N	Rank	Selector	Ignore
1	86.092	1+	85.085	0.431	0.020	0.989	53.107	59	MDRO	☐
2	102.050	1+	101.043	0.206	0.019	0.969	8.879	131	MDRO	☐
3	120.076	1+	119.069	0.306	0.027	0.966	7.940	127	MDRO	☐
4	129.094	1+	128.086	0.907	0.031	0.939	12.025	41	MDRO	☐
5	130.082	1+	129.074	0.315	0.034	0.935	3.277	171	MDRO	☐
6	147.103	1+	146.096	1.557	0.037	0.994	178.803	24	MDRO	☐
7	159.064	1+	158.057	0.393	0.032	0.973	11.383	136	MDRO	☐
8	171.101	1+	170.094	1.397	0.036	0.993	160.588	30	MDRO	☐
9	183.101	1+	182.094	2.880	0.036	0.989	221.668	17	MDRO	☐
10	187.056	1+	186.049	1.345	0.041	0.992	108.540	39	MDRO	☐
11	200.126	1+	199.119	1.193	0.062	0.929	9.515	75	MDRO	☐
12	214.111	1+	213.103	1.377	0.053	0.963	26.100	53	MDRO	☐
13	216.117	1+	215.110	1.501	0.037	0.983	62.999	40	MDRO	☐
14	223.101	1+	222.094	0.921	0.063	0.951	14.322	133	MDRO	☐
15	228.120	1+	227.113	2.955	0.047	0.808	82.788	22	MDRO	☐
16	230.056	1+	229.049	2.529	0.055	0.986	88.920	27	MDRO	☐
17	234.123	1+	233.116	6.975	0.052	0.994	652.477	8	MDRO	☐
18	239.163	1+	238.156	0.449	0.024	0.929	5.847	119	MDRO	☐
19	245.060	1+	244.053	1.191	0.072	0.897	7.524	118	MDRO	☐
20	255.117	1+	254.109	0.608	0.047	0.916	4.126	167	MDRO	☐
21	258.094	1+	257.086	0.590	0.032	0.938	8.327	108	MDRO	☐
22	261.128	1+	260.121	0.674	0.088	0.936	3.231	231	MDRO	☐
23	272.139	1+	271.131	0.354	0.032	0.944	4.648	213	MDRO	☐
24	276.117	1+	275.109	0.724	0.046	0.961	18.985	142	MDRO	☐
25	284.072	1+	283.064	1.270	0.075	0.887	4.274	95	MDRO	☐
26	287.133	1+	286.126	0.614	0.068	0.968	7.540	241	MDRO	☐
27	296.141	1+	295.134	0.854	0.092	0.885	2.501	266	MDRO	☐

Click on a peak list node on the explorer tree and you get a peak list window containing a grid of these values for all the peaks. This is where you could edit or delete a peak, if you really wanted .to. Clicking on the column headers sorts the table on that column



Having created a peak list, you might just want to save it to a file. The supported formats are Mascot Generic fomat, Comprehensive, which includes all of the peak information that we were just looking at, mzData, the XML format from the Proteomics Standards Initiative, and DTA format, which is used by Sequest.



Problems with conventional peak detection

- Failure to pick low intensity peaks
- Picking peaks that are just noise
- Selection of the wrong peak(s) in an isotopic cluster
- Need to continually ‘tweak’ parameters.

MASCOT Topic: *Mascot Distiller*

© 2009-2023 Matrix Science

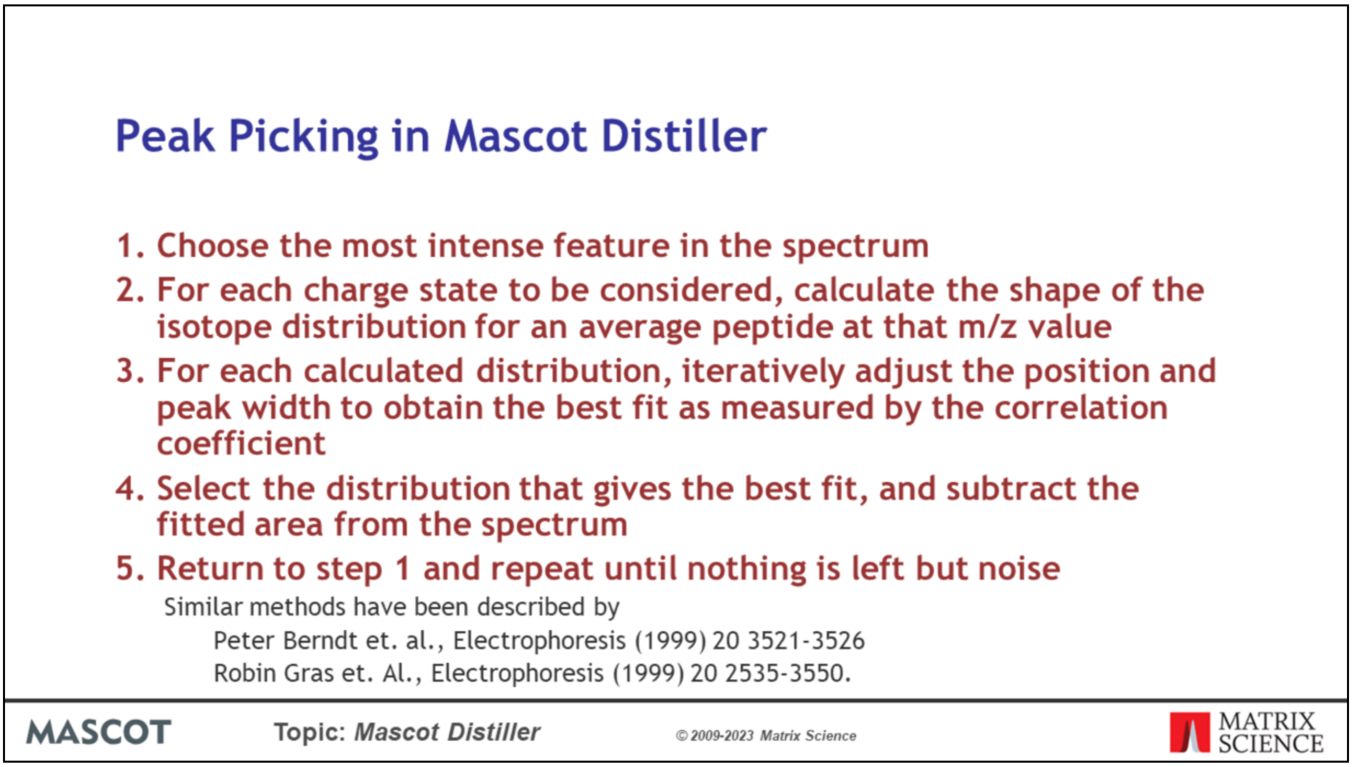
MATRIX SCIENCE

So, what is different about peak picking in Mascot Distiller?

Conventional peak detection works by smoothing the spectrum then looking for a rising gradient, which is the onset of a peak, and a falling gradient at the tail of a peak. Trouble is, this only works well if various parameters are set just right for the particular spectrum. If these parameters are not right, then we see either failure to pick low intensity peaks or picking of peaks that are just baseline noise.

Another common problem is selecting the 13C peak instead of the smaller 12C peak, so that the mass is out by a full Dalton.

The need to continually ‘tweak’ parameters is a big headache if you want to process files automatically, without looking at each spectrum.



Peak Picking in Mascot Distiller

1. Choose the most intense feature in the spectrum
2. For each charge state to be considered, calculate the shape of the isotope distribution for an average peptide at that m/z value
3. For each calculated distribution, iteratively adjust the position and peak width to obtain the best fit as measured by the correlation coefficient
4. Select the distribution that gives the best fit, and subtract the fitted area from the spectrum
5. Return to step 1 and repeat until nothing is left but noise

Similar methods have been described by

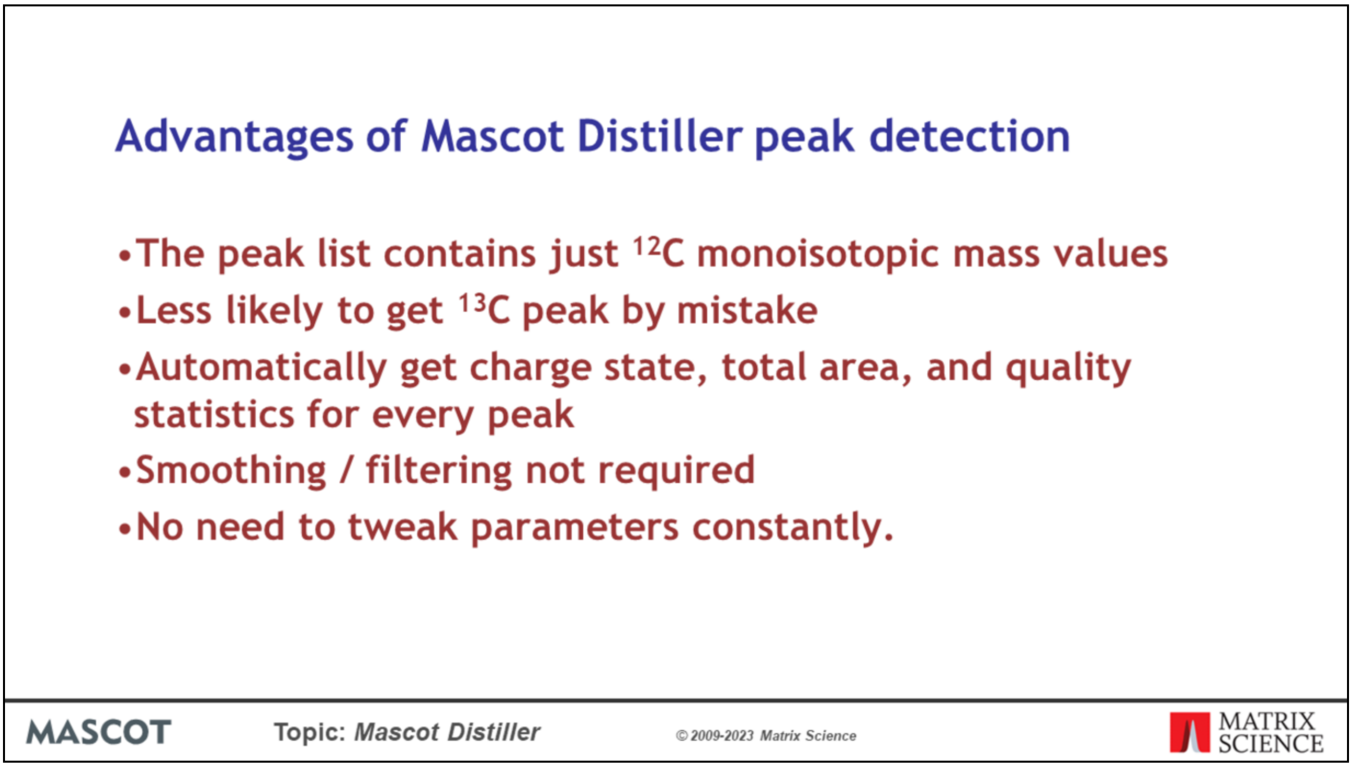
Peter Berndt et. al., Electrophoresis (1999) 20 3521-3526
Robin Gras et. Al., Electrophoresis (1999) 20 2535-3550.

MASCOT Topic: *Mascot Distiller*

© 2009-2023 Matrix Science

MATRIX SCIENCE

Mascot Distiller detects peaks by attempting to fit an ideal isotopic distribution to the experimental data. These are the steps in this process.

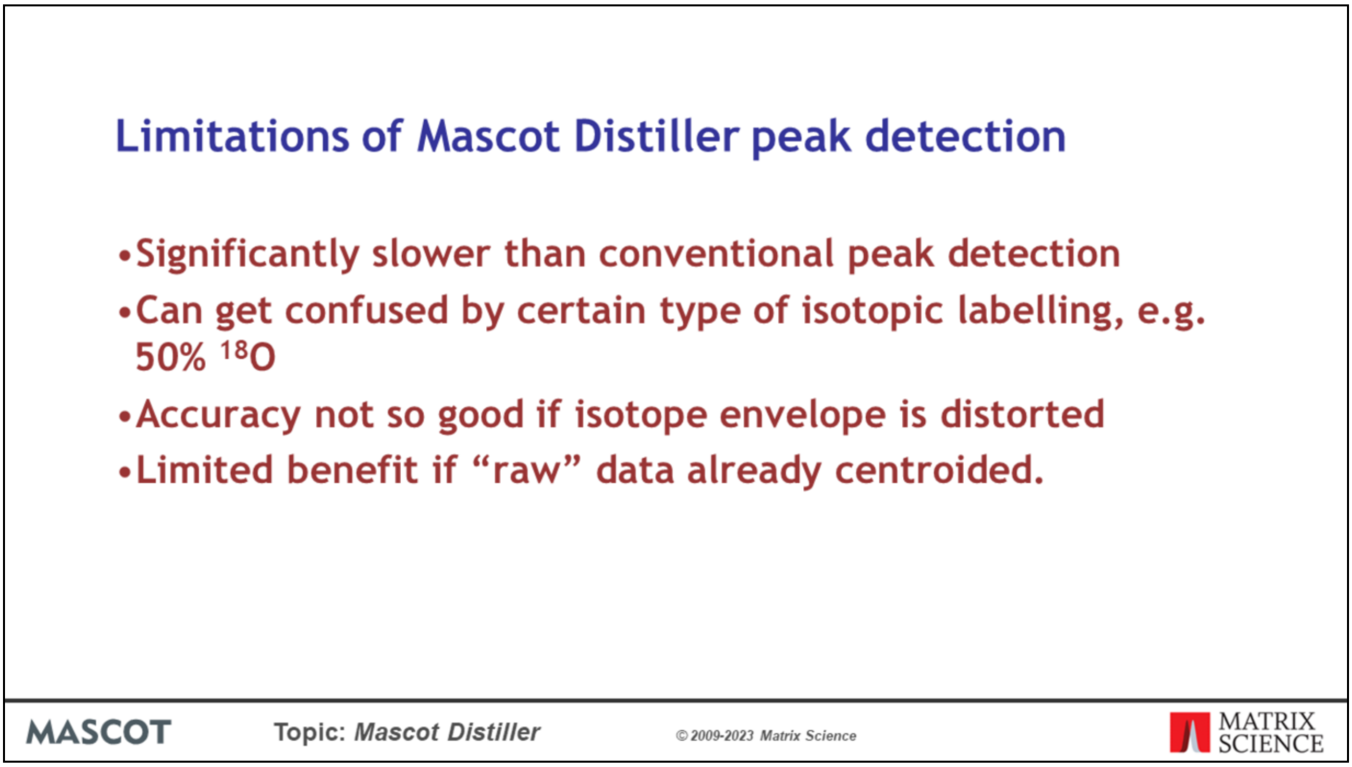


Advantages of Mascot Distiller peak detection

- The peak list contains just ^{12}C monoisotopic mass values
- Less likely to get ^{13}C peak by mistake
- Automatically get charge state, total area, and quality statistics for every peak
- Smoothing / filtering not required
- No need to tweak parameters constantly.

© 2009-2023 Matrix Science

And, these are the advantages.

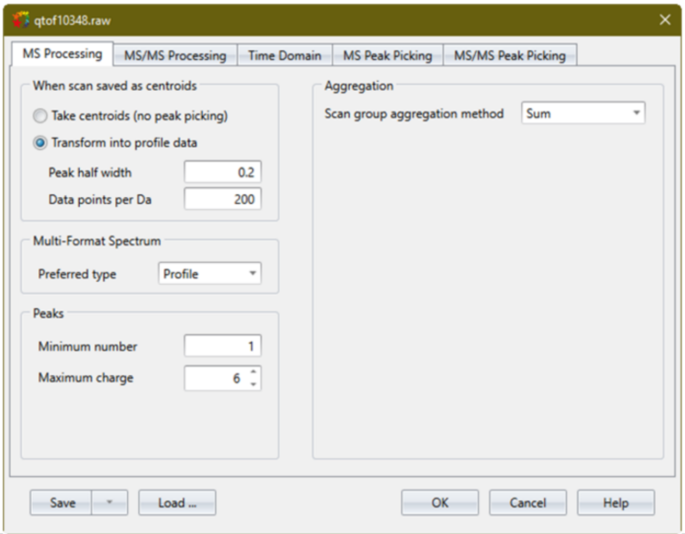


Limitations of Mascot Distiller peak detection

- Significantly slower than conventional peak detection
- Can get confused by certain type of isotopic labelling, e.g. 50% ^{18}O
- Accuracy not so good if isotope envelope is distorted
- Limited benefit if “raw” data already centroided.

Of course, nothing is perfect. Here are some of the weaknesses.

Processing options: MS



qtof10348.raw

MS Processing | MS/MS Processing | Time Domain | MS Peak Picking | MS/MS Peak Picking

When scan saved as centroids

○ Take centroids (no peak picking)

● Transform into profile data

Peak half width: 0.2

Data points per Da: 200

Multi-Format Spectrum

Preferred type: Profile

Peaks

Minimum number: 1

Maximum charge: 6

Aggregation

Scan group aggregation method: Sum

Although you don’t have to tweak the peak picking parameters constantly, this doesn’t mean there aren’t any. The point is that you can set the parameters once, for a particular instrument, and then leave them alone, confident that the peak detection will be consistently acceptable.

Sets of peak picking parameters are stored in XML text files, and can be viewed and modified using the processing options dialog. This has five tabs, and we don’t have time in this presentation to go into detail. Just press F1 when this dialog is displayed in Distiller and the on-line help will open up with complete information.

The first tab deals with MS (or survey) scans.

Mascot Distiller works best on profile data, and there is an unavoidable loss of information when data have been badly centroided. However, converting centroided data back to profile data and re-processing may yield some improvement. Mainly, the de-isotoping to reliable ^{12}C values.

If the raw scan has been saved as centroids, and **Transform into profile data** is selected, **Peak half width** is used to specify the resolution to be used for reconstructing the profile data. A Gaussian peak profile is used, and resolution is defined as the peak full width at half height in Daltons. **Data points per Da** specifies the data point spacing. This value should be at least 2 divided by the value of **Peak half Width**, to ensure that there are sufficient data points to define the peak shape accurately. If **Take centroids** is selected, Distiller will take the existing peak list from the raw file. The great advantage of this is speed. Some file formats include both profile data and centroids, in which case you can specify which is your first choice using the **preferred type** drop-down list. For example, if you want a peak list as quickly as possible, set this to centroided and select **Take centroids**. If you want the benefits of Distiller peak picking when profile data are available, then set **preferred type** to profile.

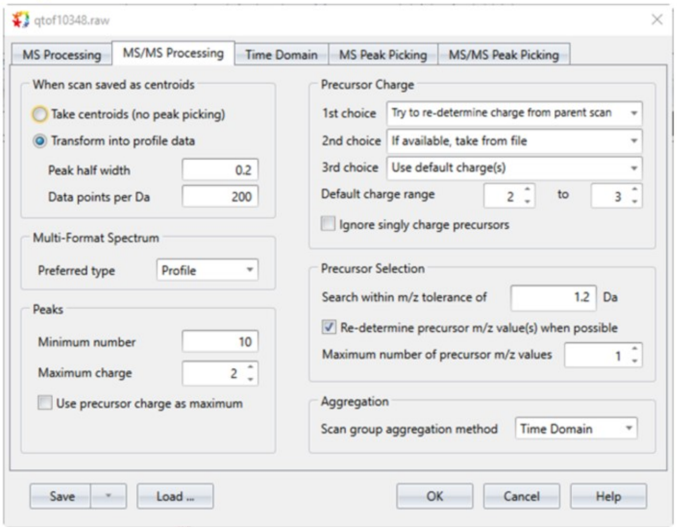
The **Aggregation Method** setting on the MS Processing tab will be used for files containing multiple MS scans, but not for time domain processing of LC-MS/MS datasets. The choices are None and Sum. When None is chosen, a separate peak list will be created for each MS scan. When Sum is chosen, all the scans will be summed prior to peak detection.

If a data file contains multiple scans from multiple samples, which are to be summed as independent groups, this must be done interactively by dragging out each group on the reconstructed TIC and choosing **Create Summed Spectrum** or **Process Range** from the Analysis menu.

Maximum charge state should be chosen carefully. Processing time increases in proportion to this number. More importantly, you should only look for high charge states in data with adequate mass resolution. If the instrument resolution means that all charge states above (say) 3 are unresolved, then it is impossible to determine from the gross width of a peak whether a peak is charge state 6 or 7. Charge state can only be reliably determined when there is some resolution between isotope peaks. Baseline resolution is not needed, just enough genuine 'ripples' to indicate that the instrumental broadening is small compared with the gross width of the isotope cluster.

Any peak list that contains less than the **Minimum number** of peaks will be discarded. If the file contains LC-MS/MS, it is important to set this value to 1, because a survey scan with only one decent peak is still important. Conversely, when processing an MS data file for peptide mass fingerprinting, this value should be set higher, because it is unlikely that you will want to search a spectrum that has less than (say) 10 peaks.

Processing options: MS/MS



When scan saved as centroids

○ Take centroids (no peak picking)

● Transform into profile data

Peak half width: 0.2

Data points per Da: 200

Multi-Format Spectrum

Preferred type: Profile

Peaks

Minimum number: 10

Maximum charge: 2

□ Use precursor charge as maximum

Precursor Charge

1st choice: Try to re-determine charge from parent scan

2nd choice: If available, take from file

3rd choice: Use default charge(s)

Default charge range: 2 to 3

□ Ignore singly charge precursors

Precursor Selection

Search within m/z tolerance of 1.2 Da

☑ Re-determine precursor m/z value(s) when possible

Maximum number of precursor m/z values: 1

Aggregation

Scan group aggregation method: Time Domain

MS/MS Processing parameters apply to all MS/MS scans, whether a single scan, a series of scans, or MS/MS scans within an LC-MS/MS dataset.

Some controls are identical to those on the MS Processing tab. These controls have been duplicated on both tabs because instruments may have different resolution capabilities for MS and MS/MS.

The aggregation method choices for MS/MS are None and Time Domain.

None is only useful for files containing a series of MS/MS scans; it is not applicable to structured LC-MS/MS data. When None is chosen, a separate peak list will be created for each MS/MS scan.

Time domain is invariably the correct choice for DDA LC-MS/MS data. Time domain means that precursor mass and charge information can be derived from survey scans, and that MS/MS scans from a common precursor should be summed together according to the rules on the Time domain tab.

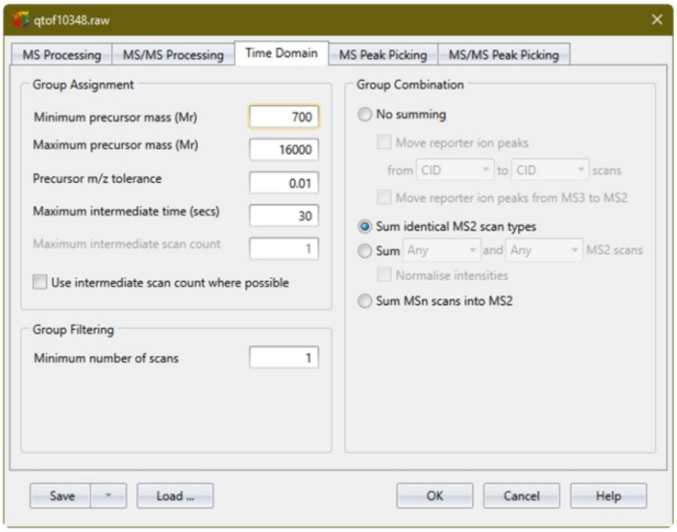
The Precursor charge frame describes the most common decision paths for assigning precursor charge state. Charge defaults are used when it is not possible to determine the precursor charge state

If ‘Ignore singly charged precursors’ is checked, spectra from singly charged precursors will be discarded. This is useful for electrospray analysis of tryptic peptides, where singly charged precursors are often noise or non-peptide contaminants.

The choices for precursor m/z can be simpler, because it is not possible to have a default m/z. The precursor m/z tolerance setting determines the maximum difference allowed between the precursor m/z value in the file and that re-determined by Distiller. The most common problem with precursor m/z is that the instrument data system has taken the 13C peak, so the precursor mass can easily be out by 1 Da. If the survey scans are high resolution, there may be multiple potential precursors in the window selected for MS/MS. If your Mascot Server is 2.5 or later, you can specify a maximum number of the most intense precursors to be picked. This can give multiple matches to chimeric spectra and is very useful when the bulk of the peaks in the MS/MS spectrum actually come from a precursor that is not the most intense in the tolerance window.

Maximum Charge state for fragments can be specified or set to the precursor charge. For the purposes of creating a peak list for a Mascot search, the correct setting depends on whether you are outputting MS/MS peaks as m/z or MH+ values, (Peak List Format tab in the Preferences dialog). A conventional peak list contains m/z values, and the maximum charge state that Mascot looks for is 2+. Hence, there is no point in spending time looking for higher charge states. However, if your data definitely includes fragment ions with higher charge states, you should choose to output fragment ions as MH+ values, and check Use precursor charge as maximum.

Processing options: Time domain



Group Assignment

Minimum precursor mass (Mr): 700

Maximum precursor mass (Mr): 16000

Precursor m/z tolerance: 0.01

Maximum intermediate time (secs): 30

Maximum intermediate scan count: 1

□ Use intermediate scan count where possible

Group Filtering

Minimum number of scans: 1

Group Combination

○ No summing

□ Move reporter ion peaks

from CID to CID scans

□ Move reporter ion peaks from MS3 to MS2

● Sum identical MS2 scan types

○ Sum Any and Any MS2 scans

□ Normalise intensities

○ Sum MSn scans into MS2

For LC-MS/MS data, the usual setting for Multi-scan data will be to use Time domain processing. The parameters on the Time domain tab then control how MS/MS scans will be summed and filtered. These settings are all fairly conventional.

Spectra from very small peptides have no value in database searching, because such short sequences can be expected to occur by chance in a large database. A setting of 700 Da for Minimum precursor mass will generally be appropriate.

The upper limit on a precursor mass in Mascot is 16000 Da, so there is little point in adding larger peptides to the peak list.

Both Precursor m/z tolerance for grouping and Max number of intermediate scans are critical parameters for accurate time domain processing. Precursor m/z tolerance for grouping requires a good estimate of how the mass spectrometer mass precision might drift during a run. Don't set this parameter too wide, because you may end up averaging together spectra from different precursors.

Maximum intermediate time or Maximum intermediate scan count require an estimate of the quality of the chromatography. Might you expect to see the same peptide elute over a period of 10 seconds, or 10 minutes, or what?

Note that scans in this context refers to survey scans. In a MassLynx file, function 1 might be the survey scan while functions 2, 3 and 4 might be MS/MS scans. Only the function 1 scans would count towards the total number of intermediate scans allowed to divide a potential group of MS/MS scans. Some knowledge of how the acquisition method was configured will be helpful in choosing an appropriate value.

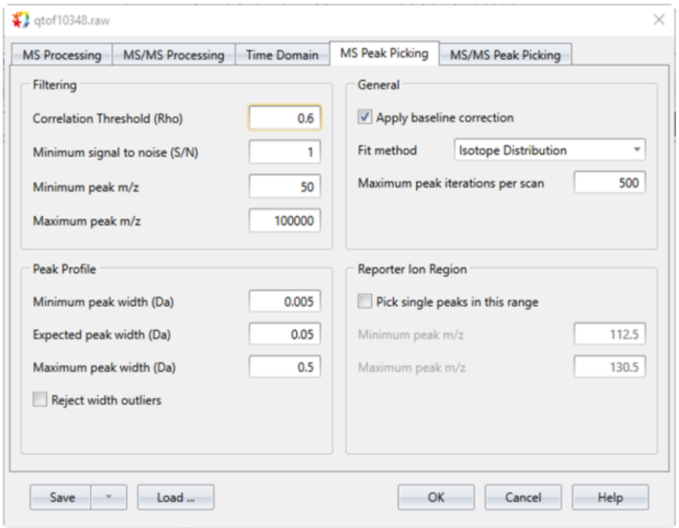
If Use intermediate scan count where possible is cleared, the Maximum intermediate scan count is disabled. If there is more than one raw file in the Distiller project, then Maximum intermediate scan count is ignored even if the checkbox is checked, because scan count has no meaning from one file to another.

Setting minimum number of scans in group requires some knowledge of how the acquisition method was configured. If the method was attempting to acquire 8 scans from each precursor, then having a single scan in the file for a particular precursor could be a good indication that the scan was triggered on a noise spike. If this is not the case, then setting this value to more than 1 could be dangerous, because valid spectra might be discarded.

Although Masslynx usually acquires multiple MS/MS scans off each precursor, this is not true for other instruments. If you are using an acquisition method with an exclusion list and only expect a single spectrum from each precursor, you will want to suppress grouping. To do this, clear Use intermediate scan count, set Maximum intermediate time to 0 and set minimum number of scans to 1

Collapse MSn scans into precursor causes the peaks in (say) an MS3 scan to be added into the peak list of the parent MS2 scan. This seems to be the best way to make use of MSn data in a database search.

Processing options: Peak picking



Filtering

Correlation Threshold (Rho) 0.6

Minimum signal to noise (S/N) 1

Minimum peak m/z 50

Maximum peak m/z 100000

General

☑ Apply baseline correction

Fit method Isotope Distribution

Maximum peak iterations per scan 500

Peak Profile

Minimum peak width (Da) 0.005

Expected peak width (Da) 0.05

Maximum peak width (Da) 0.5

□ Reject width outliers

Reporter Ion Region

□ Pick single peaks in this range

Minimum peak m/z 112.5

Maximum peak m/z 130.5

There are separate tabs for MS and MS/MS peak picking parameters, with almost identical controls. This is useful for a hybrid instrument, where the characteristics of the two scan types may be very different. If the same settings can be used for both, there is a checkbox on the MS/MS peak picking tab for ‘Same as MS Peak Picking’.

The peak picking code performs a least squares fit between a calculated isotope cluster and each candidate peak, returning the correlation coefficient, which is a measure of the similarity between the peak shapes. Ranges from 1 for perfect correlation to 0 for no correlation. Good fits to strong peaks will normally give correlation coefficients of 0.95 or better. Weak peaks will generally give lower correlation coefficients, and a cut-off of 0.7 seems to work well for all types of data.

S/N stands for signal to noise ratio. Minimum S/N needs to be set empirically for each type of instrument, though not for each dataset. Something between 1 and 10 will usually be appropriate. This is the parameter to set using some typical spectra.

Minimum and maximum peak m/z are system limits that should be set well outside the range expected for useful data.

The peak width settings are coarse ones, and do not need fine tuning. Minimum Peak Width and Maximum Peak Width are safe, conservative limits reflecting the physics of the instrument. Expected Peak Width is a starting point for iteration. The peak width will start here and can potentially go to either limit in the search for the best fit to a peak.

When Reject width outliers is checked, a robust non-parametric routine is applied to detect and remove peak width outliers. It can improve the quality of the peak list in some circumstances.

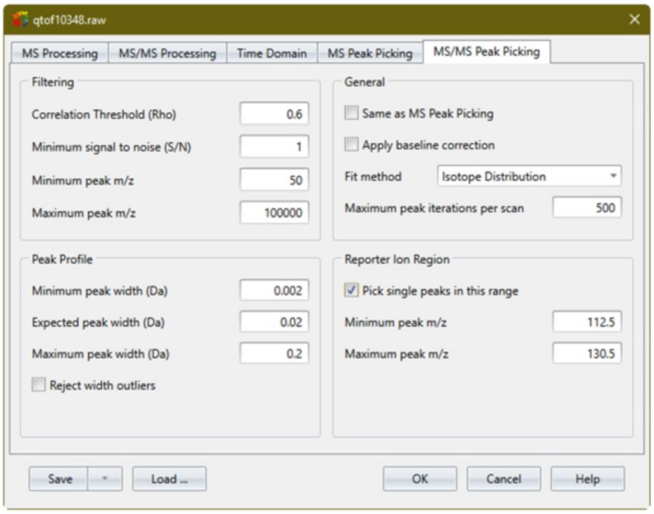
Baseline correction should be checked if the spectra have a significant baseline 'lift-off'. Generally, this will be true for MALDI-MS, but not for PSD, LCQ, QTOF, etc. If peak detection is performed on a trace with an elevated baseline, the peak list may be full of weak broad peaks, representing the signal 'under the baseline'.

Tip: The 'safe' setting is to have Baseline correction checked. It wastes a little processing time, but avoids the possibility of generating a bad peak list from a trace with an elevated baseline.

Sometimes, the peaks in a spectrum cannot be modelled using averagine. In such cases, Distiller can pick single peaks, rather than try to fit complete isotopic distributions by choosing ‘Single peak’ as the Fit method.

Maximum iterations places an upper limit on the number of detectable peaks. This limit will rarely be reached in practice. The peak picking code iterates until this limit is reached or until no significant peaks remain in the spectrum. Hence, 500 is a safe number. Setting a smaller number may reduce the overall processing time.

Processing options: Peak picking



qtof10348.raw

Filtering

Correlation Threshold (Rho): 0.6

Minimum signal to noise (S/N): 1

Minimum peak m/z: 50

Maximum peak m/z: 100000

General

☐ Same as MS Peak Picking

☐ Apply baseline correction

Fit method: Isotope Distribution

Maximum peak iterations per scan: 500

Peak Profile

Minimum peak width (Da): 0.002

Expected peak width (Da): 0.02

Maximum peak width (Da): 0.2

☐ Reject width outliers

Reporter Ion Region

☑ Pick single peaks in this range

Minimum peak m/z: 112.5

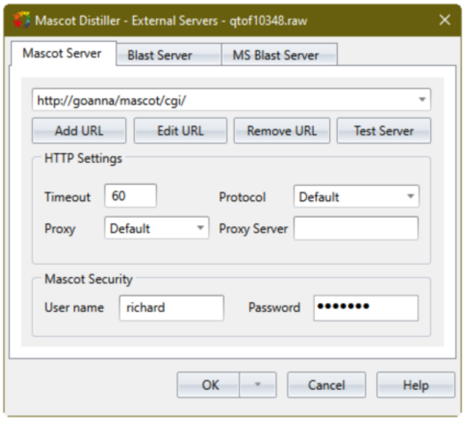
Maximum peak m/z: 130.5

© 2009-2023 Matrix Science

For an iTRAQ or TMT experiment, the ideal is to have single peak picking in the reporter ion region and isotope distribution fitting elsewhere. You can achieve this using the ‘Single peak window’ frame.

Since we don’t want this behaviour for the MS scans, the checkbox for ‘Same as MS Peak Picking’ has been cleared.

Preferences


Mascot Distiller - External Servers - qtof10348.raw

http://goanna/mascot/cgi/

HTTP Settings

Timeout 60

Protocol Default

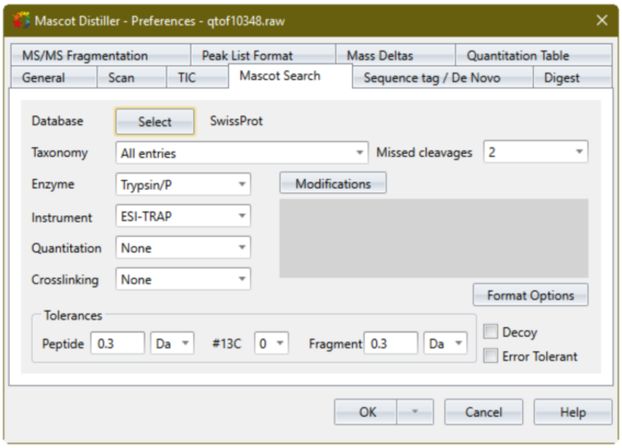
Proxy Default

Proxy Server

Mascot Security

User name richard

Password •••••••


Mascot Distiller - Preferences - qtof10348.raw

Database SwissProt

Taxonomy All entries

Missed cleavages 2

Enzyme Trypsin/P

Instrument ESI-TRAP

Quantitation None

Crosslinking None

Tolerances

Peptide 0.3 Da

#13C 0

Fragment 0.3 Da

□ Decoy

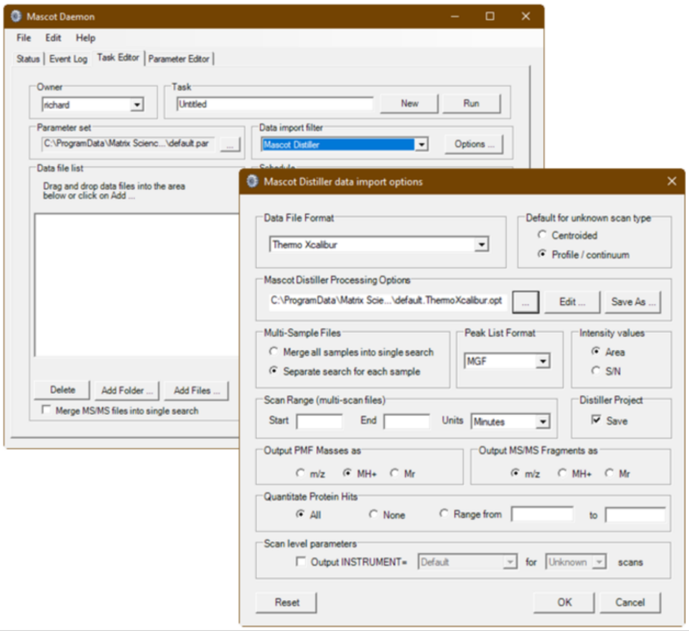
□ Error Tolerant

There are many other options dialogs in Distiller. Press F1 at any time to get context sensitive help. Two of the most important are shown here. External Servers is where you select your Mascot Server. Preferences is where you set defaults for various aspects of Distiller, such as Mascot search and *de novo*.

Automation

Mascot Distiller raw file access and peak processing is performed by a library (MDRO)

- Can be called from Mascot Daemon (Daemon Toolbox)
- Can be called from any programming language that supports Windows COM (Developer Toolbox)



Mascot Daemon

Owner: richard

Task: Untitled

Parameter set: C:\ProgramData\Matrix Scienc...\default.par

Data import filter: Mascot Distiller

Data file list: Drag and drop data files into the area below or click on Add ...

☐ Merge MS/MS files into single search

Mascot Distiller data import options

Data File Format: Thermo Xcalibur

Default for unknown scan type: ○ Centroided ● Profile / continuum

Mascot Distiller Processing Options: C:\ProgramData\Matrix Scie...\default.ThermoXcalibur.opt

Multi-Sample Files: ○ Merge all samples into single search ● Separate search for each sample

Peak List Format: MGF

Intensity values: ● Area ○ S/N

Scan Range (multi-scan files): Start End Units: Minutes

Distiller Project: ☑ Save

Output PMF Masses as: ○ m/z ● MH+ ○ Mr

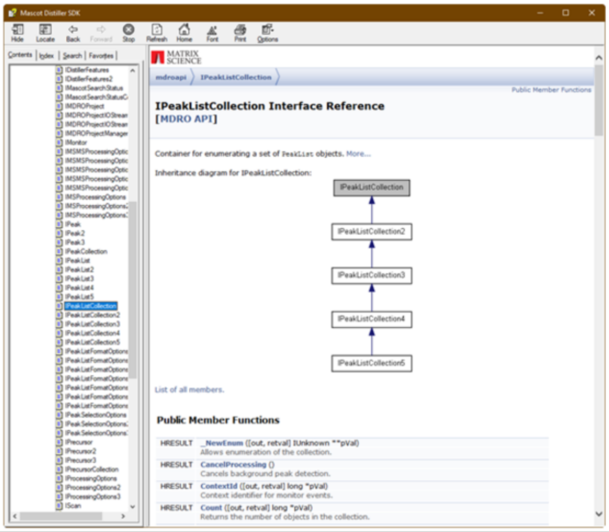
Output MS/MS Fragments as: ● m/z ○ MH+ ○ Mr

Quantitate Protein Hits: ● All ○ None ○ Range from to

Scan level parameters: ☐ Output INSTRUMENT= Default for Unknown scans

The Mascot Distiller libraries, which provides the raw file access and peak processing, can be called by other applications, such as Mascot Daemon. We'll go into more detail about how this works in a later presentation. As you can see, the Import filter options are relatively simple because all of the processing options are specified by simply selecting an options file.

Automation: Developer Toolbox



IPeakListCollection Interface Reference
[MDRO API]

Container for enumerating a set of `PeakList` objects. More...

Inheritance diagram for IPeakListCollection:

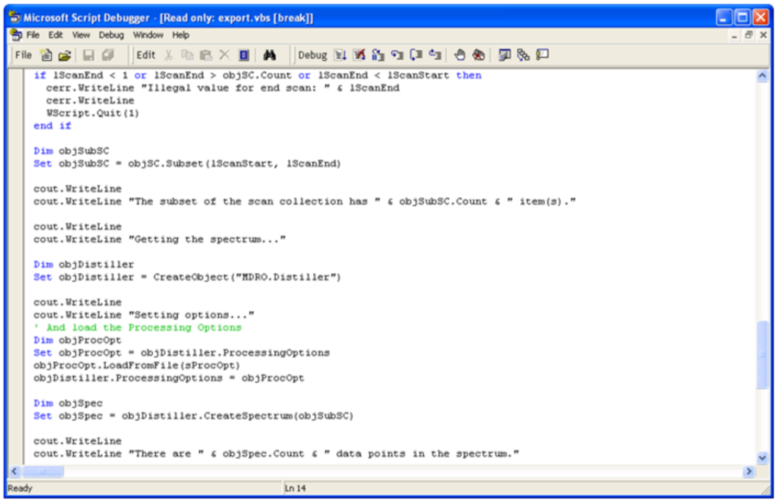
List of all members.

Public Member Functions

HRESULT	_NewEnum ([out, retval] IUnknown **pVal) Allows enumeration of the collection.
HRESULT	CancelProcessing () Cancels background peak detection.
HRESULT	ContextId ([out, retval] long *pVal) Context identifier for monitor events.
HRESULT	Count ([out, retval] long *pVal) Returns the number of objects in the collection.

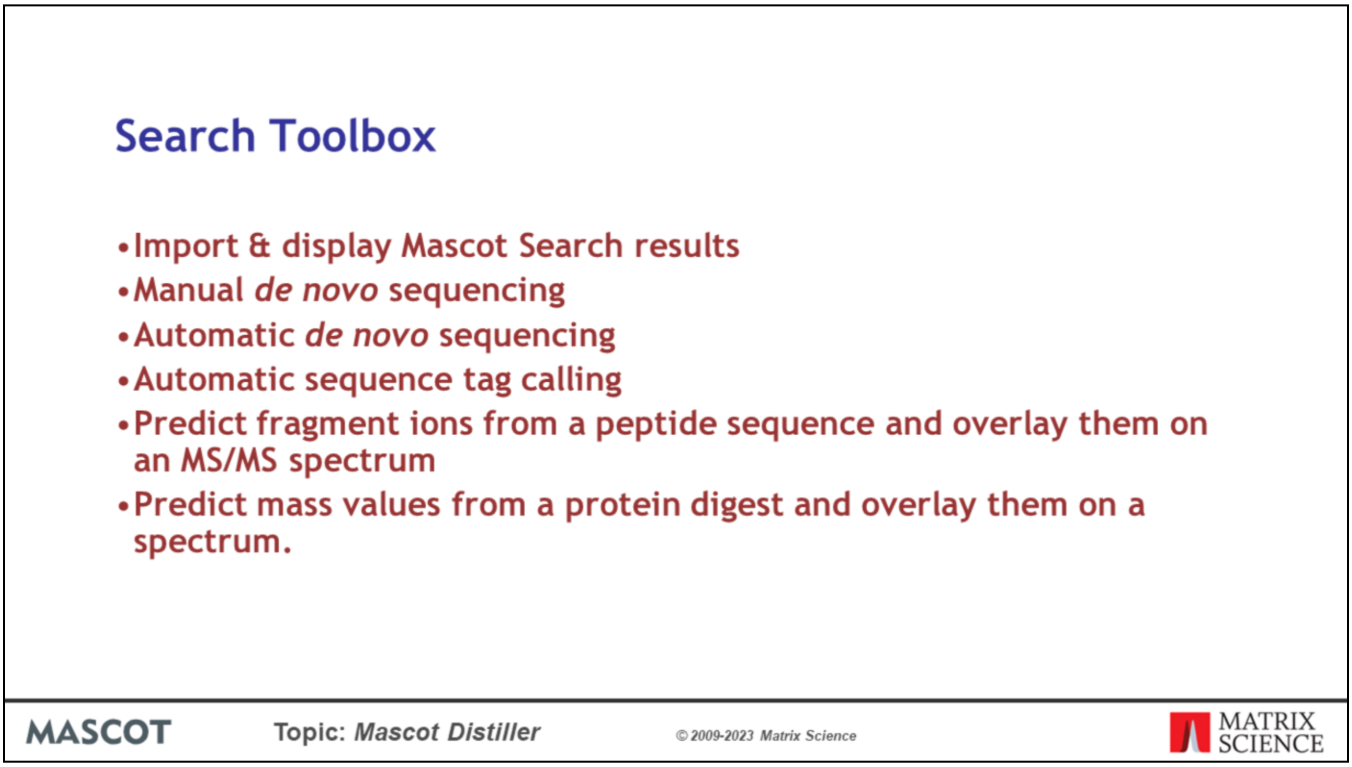
You can call the Mascot Distiller libraries from your own applications by purchasing a Developer Toolbox licence. Distiller uses COM, so can be called from most Microsoft Windows programming languages. The Developer Toolbox provides a uniform Application Programmer Interface to all of the different raw file formats, greatly reducing development time.

Automation: Developer Toolbox


Microsoft Script Debugger - [Read only: export.vbs [break]]
File Edit View Debug Window Help
File
Edit
Debug
Ready
Ln 14

```
if lScanEnd < 1 or lScanEnd > objSC.Count or lScanEnd < lScanStart then
  cerr.WriteLine "Illegal value for end scan: " & lScanEnd
  cerr.WriteLine
  WScript.Quit(1)
end if

Dim objSubSC
Set objSubSC = objSC.Subset(lScanStart, lScanEnd)

cout.WriteLine
cout.WriteLine "The subset of the scan collection has " & objSubSC.Count & " item(s)."

cout.WriteLine
cout.WriteLine "Getting the spectrum..."

Dim objDistiller
Set objDistiller = CreateObject("MDRO.Distiller")

cout.WriteLine
cout.WriteLine "Setting options..."
' And load the Processing Options
Dim objProcOpt
Set objProcOpt = objDistiller.ProcessingOptions
objProcOpt.LoadFromFile(sProcOpt)
objDistiller.ProcessingOptions = objProcOpt

Dim objSpec
Set objSpec = objDistiller.CreateSpectrum(objSubSC)

cout.WriteLine
cout.WriteLine "There are " & objSpec.Count & " data points in the spectrum."
```

Here is an example of calling Distiller from VBScript, which is a standard part of Windows. The object oriented interface makes the code very clean and simple.



Search Toolbox

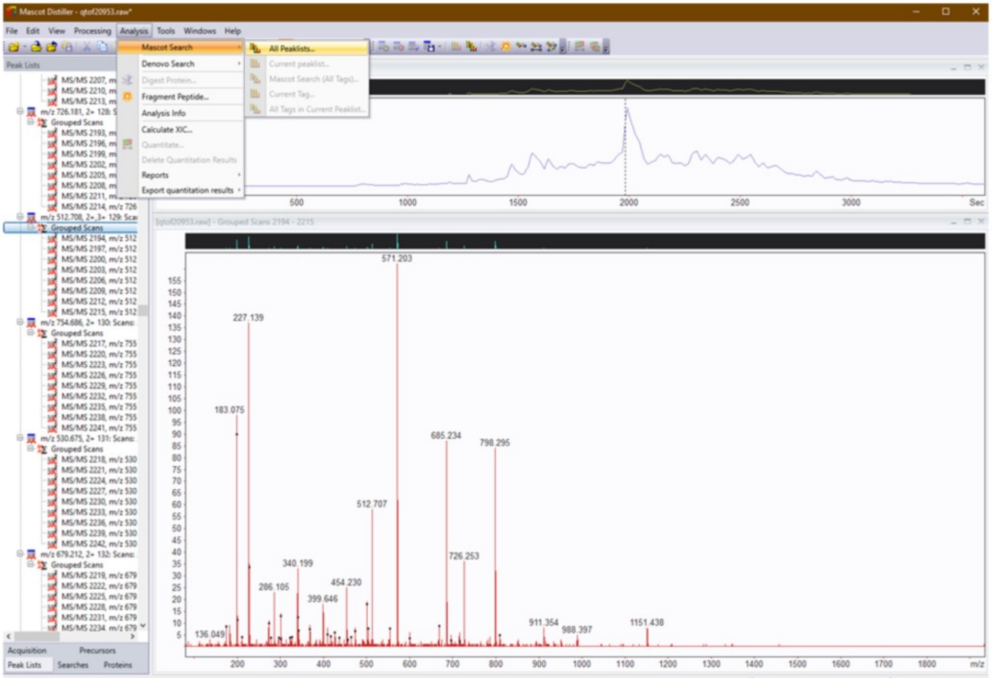
- Import & display Mascot Search results
- Manual *de novo* sequencing
- Automatic *de novo* sequencing
- Automatic sequence tag calling
- Predict fragment ions from a peptide sequence and overlay them on an MS/MS spectrum
- Predict mass values from a protein digest and overlay them on a spectrum.

MASCOT Topic: *Mascot Distiller*

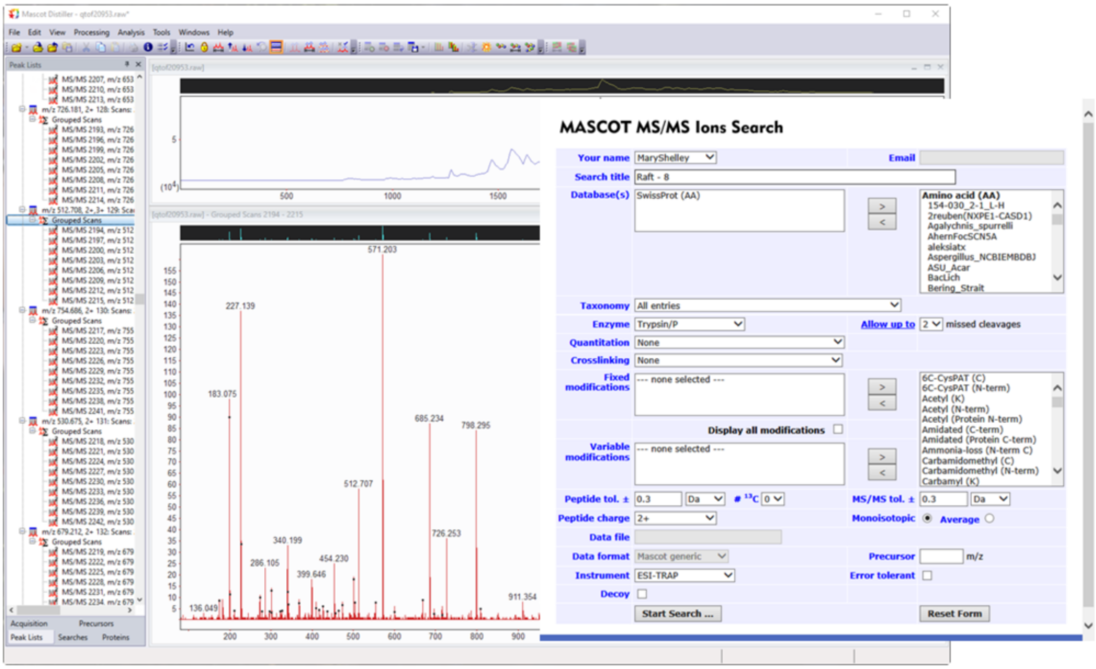
© 2009-2023 Matrix Science

MATRIX SCIENCE

The Search Toolbox is a collection of tools for protein identification and characterisation



Lets see how to submit a Mascot search from Distiller. You don’t have to save the peak list to a file. You just choose Mascot search from the context menu obtained by right clicking the top node of the peak list tree.



```
MS/MS 2207, m/z 653
MS/MS 2210, m/z 653
MS/MS 2213, m/z 653
m/z 726.181, 2+ 128: Scans:
  Grouped Scans
    MS/MS 2193, m/z 726
    MS/MS 2196, m/z 726
    MS/MS 2199, m/z 726
    MS/MS 2202, m/z 726
    MS/MS 2205, m/z 726
    MS/MS 2208, m/z 726
    MS/MS 2211, m/z 726
    MS/MS 2214, m/z 726
m/z 512.708, 2+,3+ 129: Sca
  Grouped Scans
    MS/MS 2194, m/z 512
    MS/MS 2197, m/z 512
    MS/MS 2200, m/z 512
    MS/MS 2203, m/z 512
    MS/MS 2206, m/z 512
    MS/MS 2209, m/z 512
    MS/MS 2212, m/z 512
    MS/MS 2215, m/z 512
m/z 754.686, 2+ 130: Scans:
  Grouped Scans
    MS/MS 2217, m/z 755
    MS/MS 2220, m/z 755
    MS/MS 2223, m/z 755
    MS/MS 2226, m/z 755
    MS/MS 2229, m/z 755
    MS/MS 2232, m/z 755
    MS/MS 2235, m/z 755
    MS/MS 2238, m/z 755
    MS/MS 2241, m/z 755
m/z 530.675, 2+ 131: Scans:
  Grouped Scans
    MS/MS 2218, m/z 530
    MS/MS 2221, m/z 530
    MS/MS 2224, m/z 530
    MS/MS 2227, m/z 530
    MS/MS 2230, m/z 530
    MS/MS 2233, m/z 530
    MS/MS 2236, m/z 530
    MS/MS 2239, m/z 530
    MS/MS 2242, m/z 530
m/z 679.212, 2+ 132: Scans:
  Grouped Scans
    MS/MS 2219, m/z 679
    MS/MS 2222, m/z 679
    MS/MS 2225, m/z 679
    MS/MS 2228, m/z 679
    MS/MS 2231, m/z 679
    MS/MS 2234, m/z 679
```

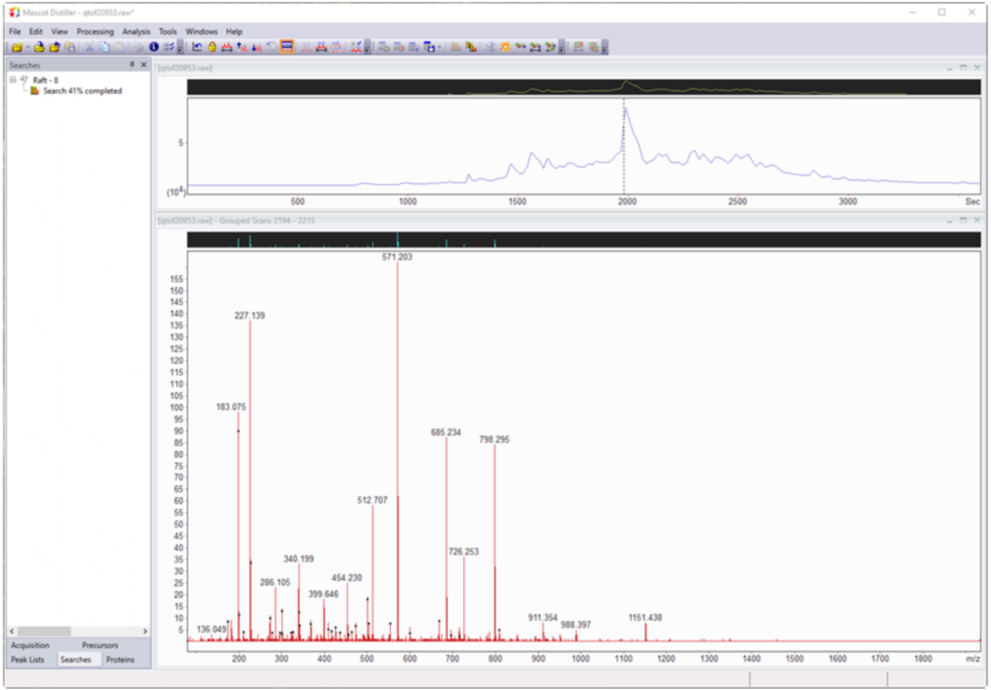
MASCOT MS/MS Ions Search

Your name: MaryShelley
Email:
Search title: Raft - 8
Database(s): SwissProt (AA)
Amino acid (AA): 154-030_2-1_L-H, 2reuben(NXPE1-CASD1), Agalychnis_spurrelli, AhernFocSCN5A, aleksiatx, Aspergillus_NCBIEMBDBJ, ASU_Acar, BacLich, Bering_Strait
Taxonomy: All entries
Enzyme: Trypsin/P
Allow up to 2 missed cleavages
Quantitation: None
Crosslinking: None
Fixed modifications: --- none selected ---
Display all modifications □
Variable modifications: --- none selected ---
6C-CysPAT (C), 6C-CysPAT (N-term), Acetyl (K), Acetyl (N-term), Acetyl (Protein N-term), Amidated (C-term), Amidated (Protein C-term), Ammonia-loss (N-term C), Carbamidomethyl (C), Carbamidomethyl (N-term), Carbamyl (K)
Peptide tol. ± 0.3 Da # ^{13}C 0
MS/MS tol. ± 0.3 Da
Peptide charge: 2+
Monoisotopic ● Average ○
Data file:
Data format: Mascot generic
Precursor: m/z
Instrument: ESI-TRAP
Error tolerant □
Decoy □

Start Search ... Reset Form

© 2009-2023 *Matrix Science*

The data are loaded into a search form and the search parameters are set to defaults specified in the Distiller project preferences. The form allows you to make any last minute changes.



MASCOT

Topic: *Mascot Distiller*

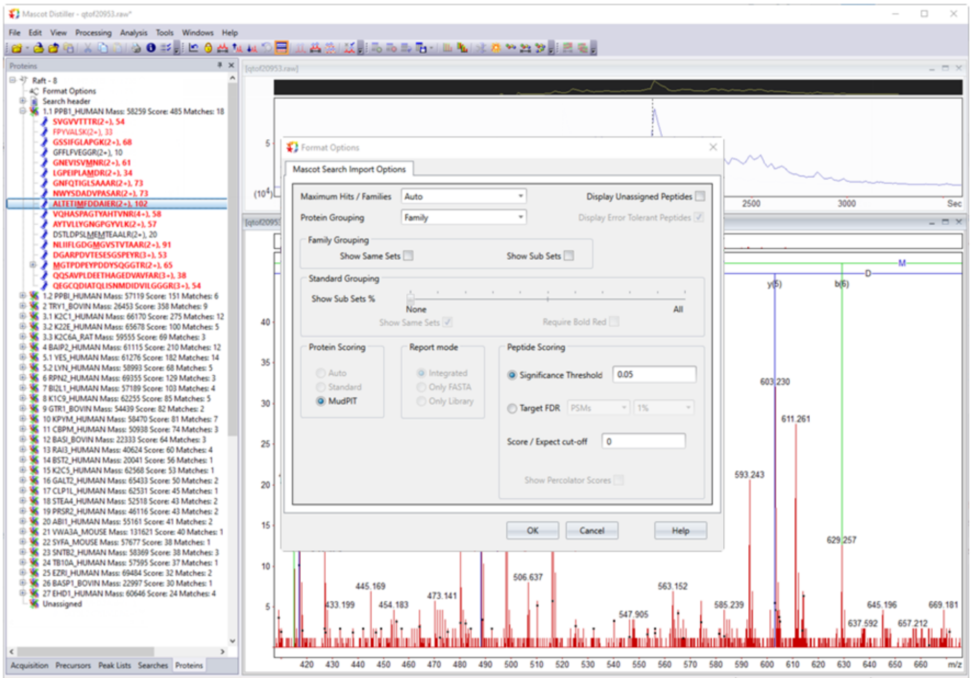
© 2009-2023 Matrix Science

MATRIX SCIENCE

If you have the search toolbox, the results are automatically retrieved from the Mascot server and displayed in DataSet Explorer. In the case of an MS/MS search, each match is labelled with the peptide sequence and Mascot peptide match score. When a peptide match node is selected, the peak assignments are displayed as sequence ladders.

It is much easier to make a judgement about the quality of the matched peaks from this kind of display than from the bitmap in the HTML result report.

The Mascot search results are actually displayed in two tabs. One is the peak lists tab, as shown here. This gives an immediate overview of all the matches to each query, and might be described as a spectrum-centric view.



Raft - 8
- Format Options
- Search header
- 1.1 PPB1_HUMAN Mass: 58259 Score: 485 Matches: 18
  - SVGVVTTTR(2+), 54
  - FPYVALSK(2+), 33
  - GSSIFGLAPGK(2+), 68
  - GFFLFVEGGR(2+), 10
  - GNEVISVMNR(2+), 61
  - LGPEIPLAMDR(2+), 34
  - GNFQTIGLSAAAR(2+), 73
  - NWYSDADVPASAR(2+), 73
  - ALTETIMFDDAIER(2+), 102
  - VQHASPAGTYAHTVNR(4+), 58
  - AYTVLLYGNGPGYVLK(2+), 57
  - DSTLDPSLMEMTEAALR(2+), 20
  - NLIIFLGDGMGVSTVTAAR(2+), 91
  - DGARPDVTESESGSPEYR(3+), 53
  - MGTPDPEYPDDYSQGGTR(2+), 65
  - QQSAVPLDEETHAGEDVAVFAR(3+), 38
  - QEGCQDIATQLISNMDIDVILGGGR(3+), 54
- 1.2 PPBI_HUMAN Mass: 57119 Score: 151 Matches: 6
- 2 TRY1_BOVIN Mass: 26453 Score: 358 Matches: 9
- 3.1 K2C1_HUMAN Mass: 66170 Score: 275 Matches: 12
- 3.2 K22E_HUMAN Mass: 65678 Score: 100 Matches: 5
- 3.3 K2C6A_RAT Mass: 59555 Score: 69 Matches: 3
- 4 BAIP2_HUMAN Mass: 61115 Score: 210 Matches: 12
- 5.1 YES_HUMAN Mass: 61276 Score: 182 Matches: 14
- 5.2 LYN_HUMAN Mass: 58993 Score: 68 Matches: 5
- 6 RPN2_HUMAN Mass: 69355 Score: 129 Matches: 3
- 7 BI2L1_HUMAN Mass: 57189 Score: 103 Matches: 4
- 8 K1C9_HUMAN Mass: 62255 Score: 85 Matches: 5
- 9 GTR1_BOVIN Mass: 54439 Score: 82 Matches: 2
- 10 KPYM_HUMAN Mass: 58470 Score: 81 Matches: 7
- 11 CBPM_HUMAN Mass: 50938 Score: 74 Matches: 3
- 12 BASI_BOVIN Mass: 22333 Score: 64 Matches: 3
- 13 RAI3_HUMAN Mass: 40624 Score: 60 Matches: 4
- 14 BST2_HUMAN Mass: 20041 Score: 56 Matches: 1
- 15 K2C5_HUMAN Mass: 62568 Score: 53 Matches: 1
- 16 GALT2_HUMAN Mass: 65433 Score: 50 Matches: 2
- 17 CLP1L_HUMAN Mass: 62531 Score: 45 Matches: 1
- 18 STEA4_HUMAN Mass: 52518 Score: 43 Matches: 2
- 19 PRSR2_HUMAN Mass: 46116 Score: 43 Matches: 2
- 20 ABI1_HUMAN Mass: 55161 Score: 41 Matches: 2
- 21 VWA3A_MOUSE Mass: 131621 Score: 40 Matches: 1
- 22 SYFA_MOUSE Mass: 57677 Score: 38 Matches: 1
- 23 SNTB2_HUMAN Mass: 58369 Score: 38 Matches: 3
- 24 TB10A_HUMAN Mass: 57595 Score: 37 Matches: 1
- 25 EZRI_HUMAN Mass: 69484 Score: 32 Matches: 2
- 26 BASP1_BOVIN Mass: 22997 Score: 30 Matches: 1
- 27 EHD1_HUMAN Mass: 60646 Score: 24 Matches: 4
- Unassigned

Format Options

Mascot Search Import Options

Maximum Hits / Families: Auto
Display Unassigned Peptides □
Protein Grouping: Family
Display Error Tolerant Peptides ☑

Family Grouping
Show Same Sets □ Show Sub Sets □

Standard Grouping
Show Sub Sets % None ... All
Show Same Sets ☑ Require Bold Red □

Protein Scoring
○ Auto
○ Standard
● MudPIT

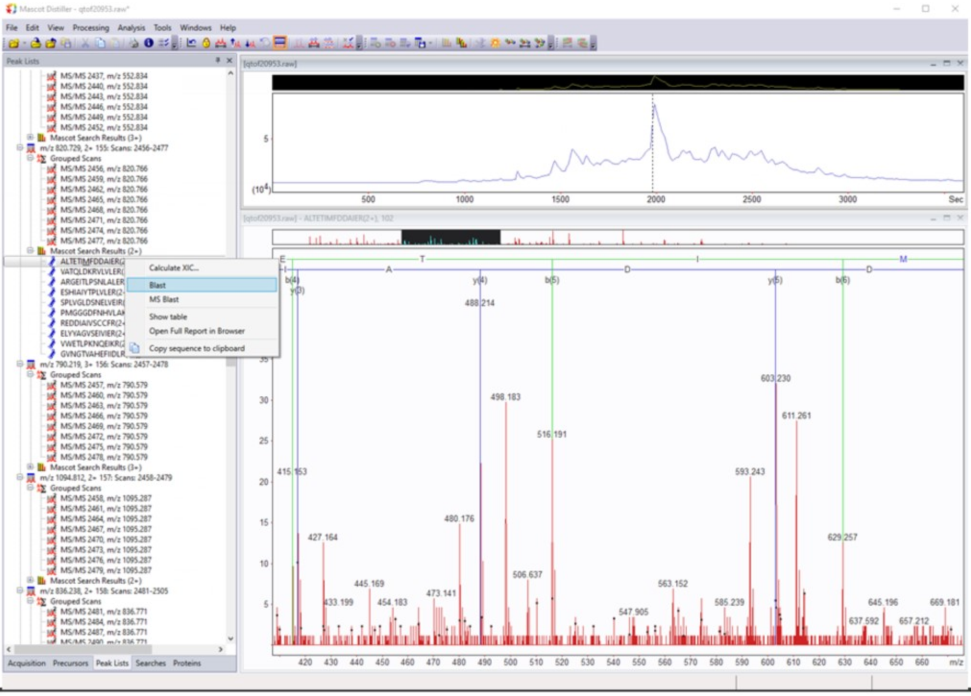
Report mode
● Integrated
○ Only FASTA
○ Only Library

Peptide Scoring
● Significance Threshold 0.05
○ Target FDR PSMs 1%
Score / Expect cut-off 0
Show Percolator Scores □

OK Cancel Help

The Proteins tab provides a protein-centric view, closely resembling the Mascot Protein Family Summary report, with the peptide matches grouped into protein hits. This is a more natural arrangement if you are only interested in the significant matches, or the matches assigned to a particular protein.

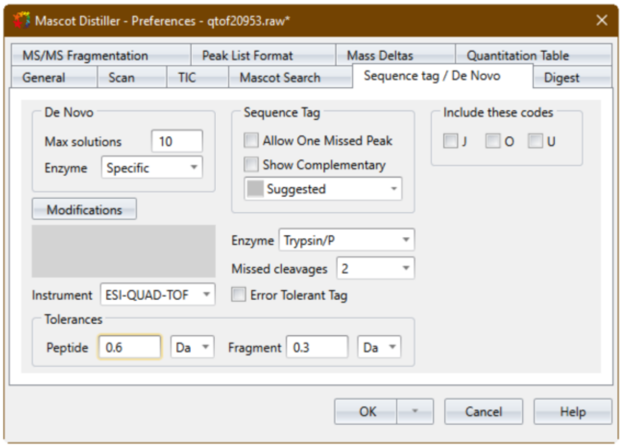
As far as possible, the selected node is synchronised between the two tabs, making it easy to switch between the two views. The Format options node at the top of the proteins tree is similar to the Format controls at the top of the HTML summary reports.



Going back to the peak lists tab, right click any match for a context menu containing relevant options, such as a link out to Blast or MS-Blast.

De novo

- Novel algorithm
- Scores approximate to Mascot scores
- Single spectrum or entire dataset



Mascot Distiller - Preferences - qtof20953.raw*

MS/MS Fragmentation | Peak List Format | Mass Deltas | Quantitation Table | General | Scan | TIC | Mascot Search | Sequence tag / De Novo | Digest

De Novo
Max solutions: 10
Enzyme: Specific
Modifications
Instrument: ESI-QUAD-TOF

Sequence Tag
□ Allow One Missed Peak
□ Show Complementary
Suggested
Enzyme: Trypsin/P
Missed cleavages: 2
□ Error Tolerant Tag

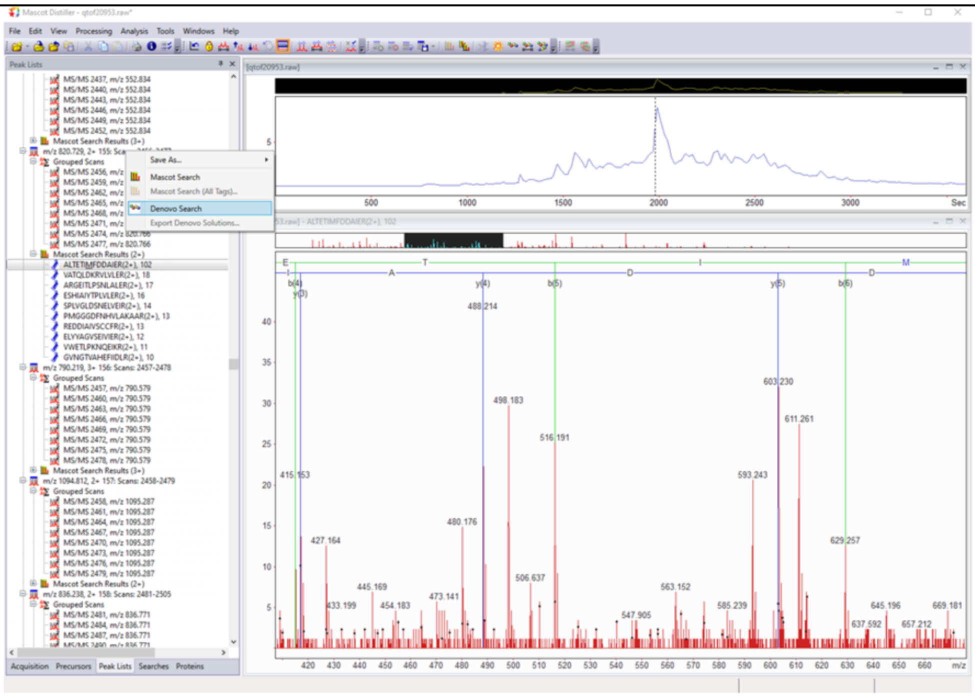
Include these codes
□ J □ O □ U

Tolerances
Peptide: 0.6 Da
Fragment: 0.3 Da

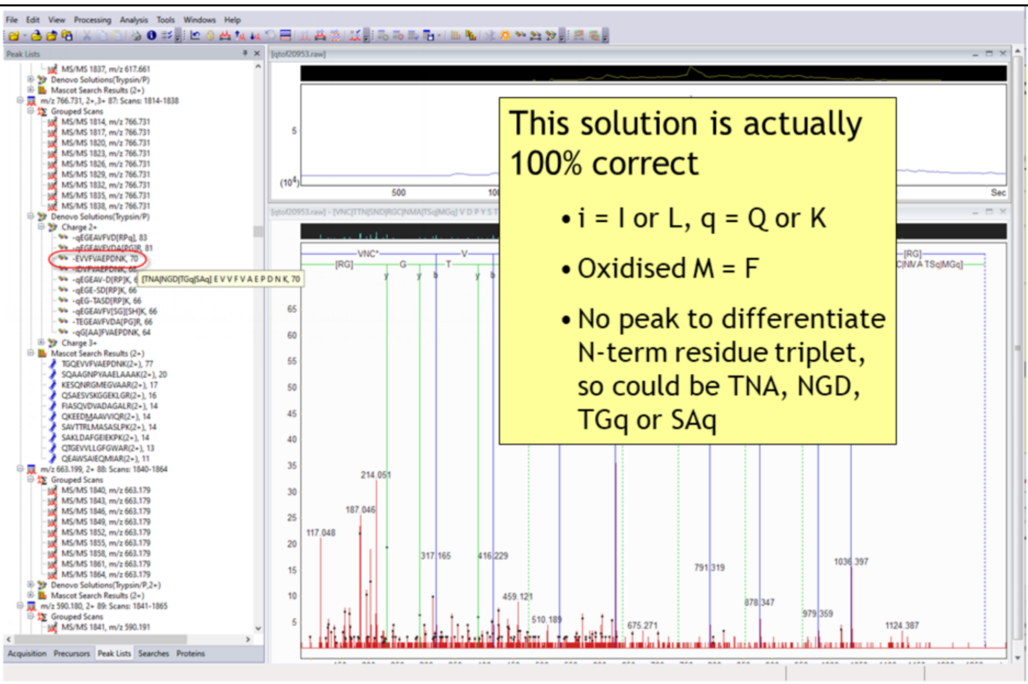
The search toolbox includes a powerful de novo sequencing module.

The score assigned to a de novo solution is similar to a Mascot score. In general, these scores will be higher than you expect to see in a Mascot database search, because the algorithm has selected the best matching sequence from all possible sequences, rather than the limited number of sequences found in any database. So, you should not judge the quality of the match by applying any rule of thumb or significance threshold to the score. However, if you get the same solution by de novo and by database search, using identical parameters, you should find the Mascot scores are very close.

You can *de novo* sequence a single spectrum or all the MS/MS spectra in the file.



The starting point can be any MS/MS scan that has been processed to create a peak list. Right click the peak list node in Dataset Explorer and choose ‘de novo Search’, or choose the de novo button from the toolbar when a Summed Scans node is selected.



This solution is actually 100% correct

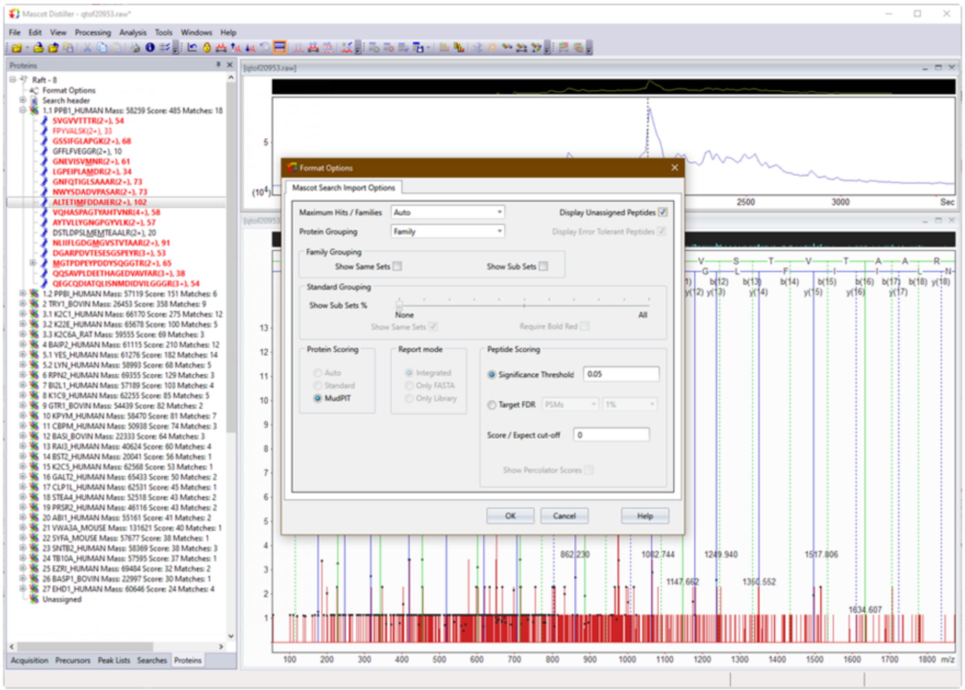
- i = I or L, q = Q or K
- Oxidised M = F
- No peak to differentiate N-term residue triplet, so could be TNA, NGD, TGq or SAq

Good signal to noise and good mass accuracy are critical for successful de novo sequencing; much more so than in database searching. GIGO (garbage in - garbage out) is guaranteed.

In a *de novo* solution, i always represents I or L, and q represent Q or K when the mass tolerance does not allow these residues to be distinguished. However, K is assumed at the C terminus of a peptide when tryptic specificity applies

Ambiguity is indicated by a dash in the sequence. The tooltip shows details of the ambiguity in square brackets, using pipe symbols to separate alternatives. Note that the order of the pairs and triplets is undefined, so the triplets TNA, NGD, TGq or Saq can each be expanded to 6 different combinations (3x2x1).

Although the example shown here looks very different to the Mascot database match, they are actually in perfectly agreement. Some uncertainty is unavoidable in *de novo*, because the search space is so very much larger



Raft - 8
Format Options
Search header
1.1 PPB1_HUMAN Mass: 58259 Score: 485 Matches: 18
SVGVVTTTR(2+), 54
FPYVALSK(2+), 33
GSSIFGLAPGK(2+), 68
GFFLFVEGGR(2+), 10
GNEVISVMNR(2+), 61
LGPEIPLAMDR(2+), 34
GNFQTIGLSAAAR(2+), 73
NWYSDADVPASAR(2+), 73
ALTETIMFDDAIER(2+), 102
VQHASPAGTYAHTVNR(4+), 58
AYTVLLYGNGPGYVLK(2+), 57
DSTLDPSLMEMTEAALR(2+), 20
NLIIFLGDGMGVSTVTAAR(2+), 91
DGARPDVTESESGSPEYR(3+), 53
MGTPDPEYPDDYSQGGTR(2+), 65
QQSAVPLDEETHAGEDVAVFAR(3+), 38
QEGCQDIATQLISNMDIDVILGGGR(3+), 54
1.2 PPBI_HUMAN Mass: 57119 Score: 151 Matches: 6
2 TRY1_BOVIN Mass: 26453 Score: 358 Matches: 9
3.1 K2C1_HUMAN Mass: 66170 Score: 275 Matches: 12
3.2 K22E_HUMAN Mass: 65678 Score: 100 Matches: 5
3.3 K2C6A_RAT Mass: 59555 Score: 69 Matches: 3
4 BAIP2_HUMAN Mass: 61115 Score: 210 Matches: 12
5.1 YES_HUMAN Mass: 61276 Score: 182 Matches: 14
5.2 LYN_HUMAN Mass: 58993 Score: 68 Matches: 5
6 RPN2_HUMAN Mass: 69355 Score: 129 Matches: 3
7 BI2L1_HUMAN Mass: 57189 Score: 103 Matches: 4
8 K1C9_HUMAN Mass: 62255 Score: 85 Matches: 5
9 GTR1_BOVIN Mass: 54439 Score: 82 Matches: 2
10 KPYM_HUMAN Mass: 58470 Score: 81 Matches: 7
11 CBPM_HUMAN Mass: 50938 Score: 74 Matches: 3
12 BASI_BOVIN Mass: 22333 Score: 64 Matches: 3
13 RAI3_HUMAN Mass: 40624 Score: 60 Matches: 4
14 BST2_HUMAN Mass: 20041 Score: 56 Matches: 1
15 K2C5_HUMAN Mass: 62568 Score: 53 Matches: 1
16 GALT2_HUMAN Mass: 65433 Score: 50 Matches: 2
17 CLP1L_HUMAN Mass: 62531 Score: 45 Matches: 1
18 STEA4_HUMAN Mass: 52518 Score: 43 Matches: 2
19 PRSR2_HUMAN Mass: 46116 Score: 43 Matches: 2
20 ABI1_HUMAN Mass: 55161 Score: 41 Matches: 2
21 VWA3A_MOUSE Mass: 131621 Score: 40 Matches: 1
22 SYFA_MOUSE Mass: 57677 Score: 38 Matches: 1
23 SNTB2_HUMAN Mass: 58369 Score: 38 Matches: 3
24 TB10A_HUMAN Mass: 57595 Score: 37 Matches: 1
25 EZRI_HUMAN Mass: 69484 Score: 32 Matches: 2
26 BASP1_BOVIN Mass: 22997 Score: 30 Matches: 1
27 EHD1_HUMAN Mass: 60646 Score: 24 Matches: 4
Unassigned

Format Options

Mascot Search Import Options

Maximum Hits / Families: Auto
Display Unassigned Peptides ☑
Protein Grouping: Family
Display Error Tolerant Peptides ☑

Family Grouping
Show Same Sets □ Show Sub Sets □

Standard Grouping
Show Sub Sets %: None … All
Show Same Sets ☑ Require Bold Red □

Protein Scoring: ○ Auto ○ Standard ● MudPIT

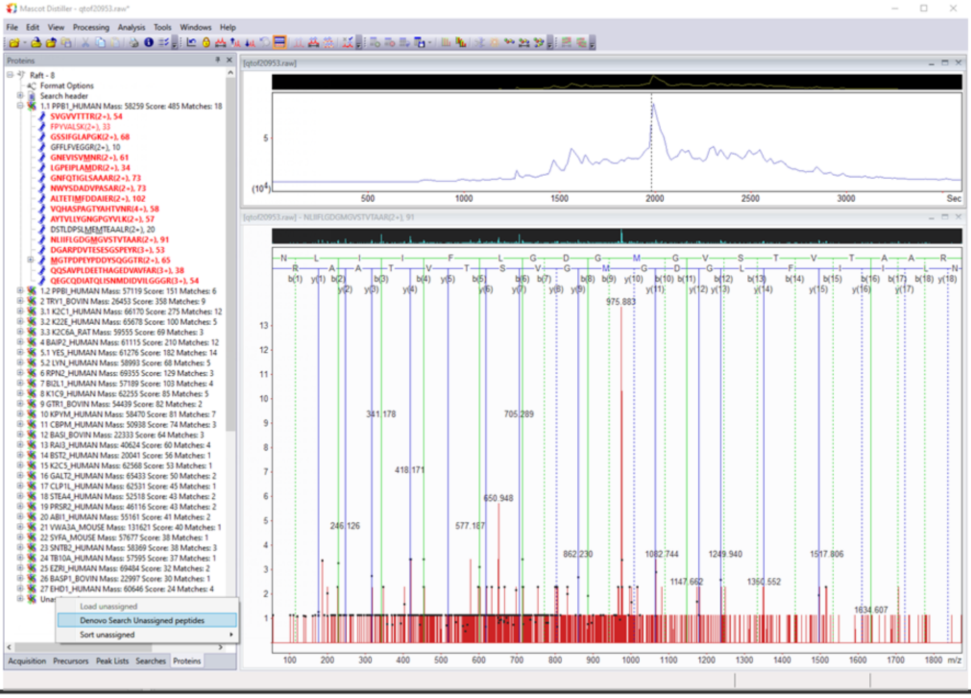
Report mode: ● Integrated ○ Only FASTA ○ Only Library

Peptide Scoring
● Significance Threshold: 0.05
○ Target FDR: PSMs 1%
Score / Expect cut-off: 0
Show Percolator Scores □

OK Cancel Help

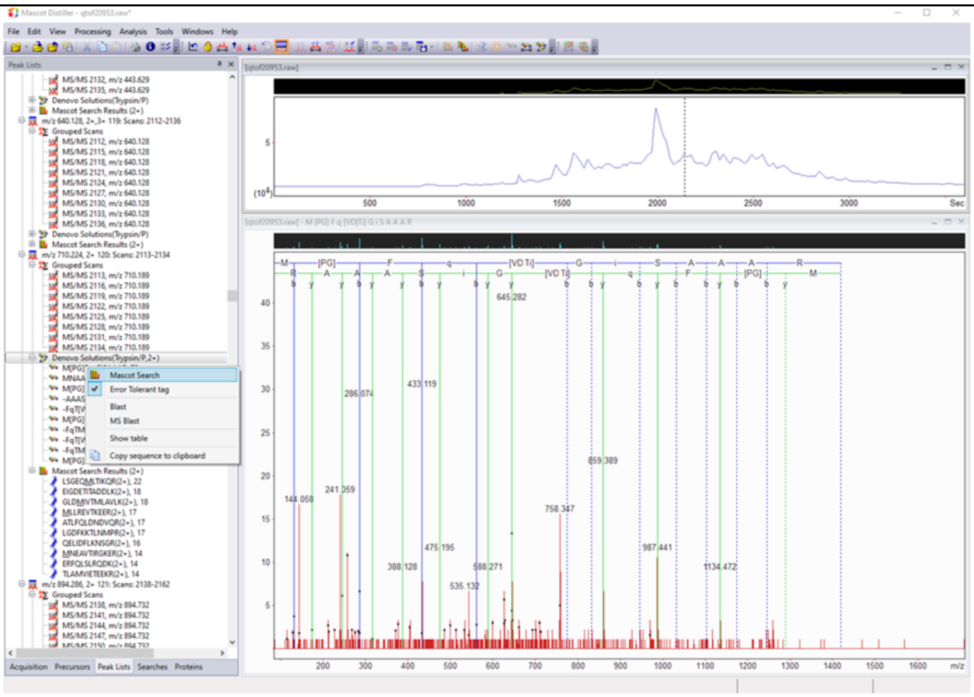
To *de novo* sequence a complete peak list collection, or the peak lists in the currently displayed TIC range, use the context menu obtained by right-clicking the root (world) node of the peak lists tree.

The most efficient way to *de novo* only those spectra that failed to give decent matches in the Mascot search is to switch to the proteins tree, click on Format Options, and choose to load the unassigned queries.


Mascot Distiller - qtof20953.raw*
File Edit View Processing Analysis Tools Windows Help
Proteins
[qtof20953.raw]
[qtof20953.raw] - NLIIFLGDGMGVSTVTAAR(2+), 91
Load unassigned
Denovo Search Unassigned peptides
Sort unassigned
Acquisition Precursors Peak Lists Searches Proteins

```
Raft - 8
  Format Options
  Search header
  1.1 PPB1_HUMAN Mass: 58259 Score: 485 Matches: 18
    SVGVVTTTR(2+), 54
    FPYVALSK(2+), 33
    GSSIFGLAPGK(2+), 68
    GFFLFVEGGR(2+), 10
    GNEVISVMNR(2+), 61
    LGPEIPLAMDR(2+), 34
    GNFQTIGLSAAAR(2+), 73
    NWYSDADVPASAR(2+), 73
    ALTETIMFDDAIER(2+), 102
    VQHASPAGTYAHTVNR(4+), 58
    AYTVLLYGNGPGYVLK(2+), 57
    DSTLDPSLMEMTEAALR(2+), 20
    NLIIFLGDGMGVSTVTAAR(2+), 91
    DGARPDVTESESGSPEYR(3+), 53
    MGTPDPEYPDDYSQGGTR(2+), 65
    QQSAVPLDEETHAGEDVAVFAR(3+), 38
    QEGCQDIATQLISNMDIDVILGGGR(3+), 54
  1.2 PPBI_HUMAN Mass: 57119 Score: 151 Matches: 6
  2 TRY1_BOVIN Mass: 26453 Score: 358 Matches: 9
  3.1 K2C1_HUMAN Mass: 66170 Score: 275 Matches: 12
  3.2 K22E_HUMAN Mass: 65678 Score: 100 Matches: 5
  3.3 K2C6A_RAT Mass: 59555 Score: 69 Matches: 3
  4 BAIP2_HUMAN Mass: 61115 Score: 210 Matches: 12
  5.1 YES_HUMAN Mass: 61276 Score: 182 Matches: 14
  5.2 LYN_HUMAN Mass: 58993 Score: 68 Matches: 5
  6 RPN2_HUMAN Mass: 69355 Score: 129 Matches: 3
  7 BI2L1_HUMAN Mass: 57189 Score: 103 Matches: 4
  8 K1C9_HUMAN Mass: 62255 Score: 85 Matches: 5
  9 GTR1_BOVIN Mass: 54439 Score: 82 Matches: 2
  10 KPYM_HUMAN Mass: 58470 Score: 81 Matches: 7
  11 CBPM_HUMAN Mass: 50938 Score: 74 Matches: 3
  12 BASI_BOVIN Mass: 22333 Score: 64 Matches: 3
  13 RAI3_HUMAN Mass: 40624 Score: 60 Matches: 4
  14 BST2_HUMAN Mass: 20041 Score: 56 Matches: 1
  15 K2C5_HUMAN Mass: 62568 Score: 53 Matches: 1
  16 GALT2_HUMAN Mass: 65433 Score: 50 Matches: 2
  17 CLP1L_HUMAN Mass: 62531 Score: 45 Matches: 1
  18 STEA4_HUMAN Mass: 52518 Score: 43 Matches: 2
  19 PRSR2_HUMAN Mass: 46116 Score: 43 Matches: 2
  20 ABI1_HUMAN Mass: 55161 Score: 41 Matches: 2
  21 VWA3A_MOUSE Mass: 131621 Score: 40 Matches: 1
  22 SYFA_MOUSE Mass: 57677 Score: 38 Matches: 1
  23 SNTB2_HUMAN Mass: 58369 Score: 38 Matches: 3
  24 TB10A_HUMAN Mass: 57595 Score: 37 Matches: 1
  25 EZRI_HUMAN Mass: 69484 Score: 32 Matches: 2
  26 BASP1_BOVIN Mass: 22997 Score: 30 Matches: 1
  27 EHD1_HUMAN Mass: 60646 Score: 24 Matches: 4
  Una
```

Use the context menu obtained by right-clicking the root node or unassigned node to *de novo* just the unassigned queries.



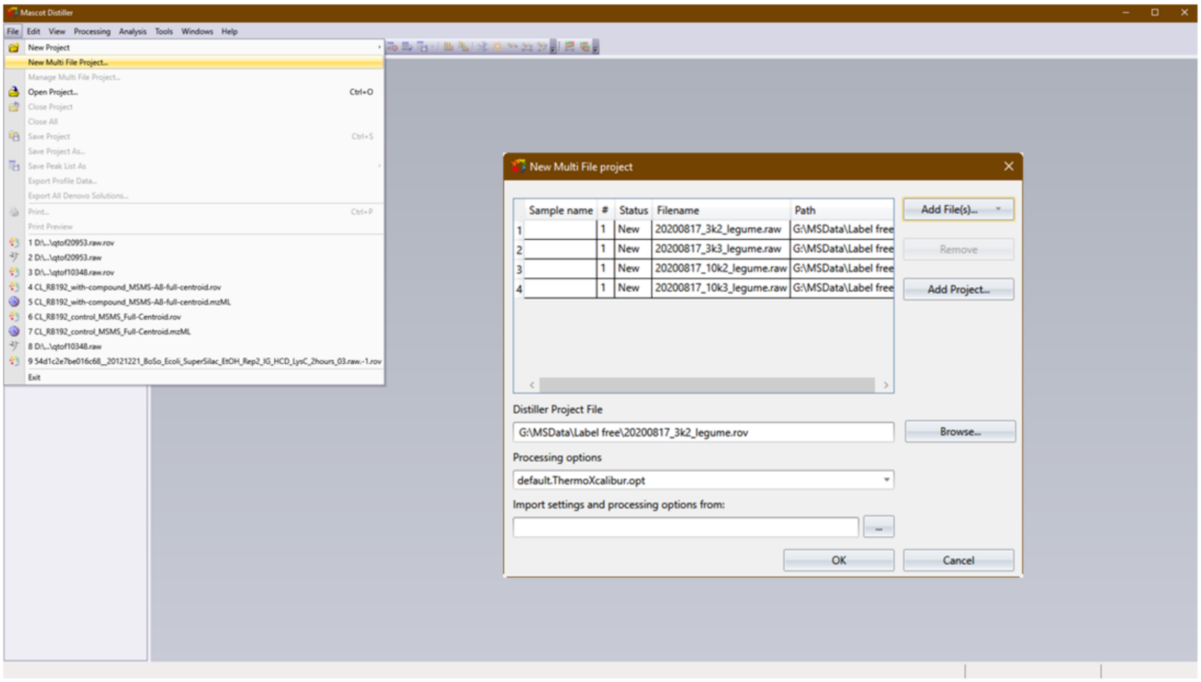
Mascot Distiller - qtof20953.raw*

Peak Lists

MS/MS 2132, m/z 443.629
MS/MS 2135, m/z 443.629
Denovo Solutions(Trypsin/P)
Mascot Search Results (2+)
m/z 640.128, 2+,3+ 119: Scans: 2112-2136
Grouped Scans
MS/MS 2112, m/z 640.128
MS/MS 2115, m/z 640.128
MS/MS 2118, m/z 640.128
MS/MS 2121, m/z 640.128
MS/MS 2124, m/z 640.128
MS/MS 2127, m/z 640.128
MS/MS 2130, m/z 640.128
MS/MS 2133, m/z 640.128
MS/MS 2136, m/z 640.128
Denovo Solutions(Trypsin/P)
Mascot Search Results (2+)
m/z 710.224, 2+ 120: Scans: 2113-2134
Grouped Scans
MS/MS 2113, m/z 710.189
MS/MS 2116, m/z 710.189
MS/MS 2119, m/z 710.189
MS/MS 2122, m/z 710.189
MS/MS 2125, m/z 710.189
MS/MS 2128, m/z 710.189
MS/MS 2131, m/z 710.189
MS/MS 2134, m/z 710.189
Denovo Solutions(Trypsin/P,2+)
M[PG]
MNAA
M[PG]
-AAAS
-FqT[V
M[PG]
-FqTM
-FqT[V
-FqTM
M[PG]
Mascot Search Results (2+)
LSGEQMLTIKQR(2+), 22
EIGDETITADDLK(2+), 18
GLDMIVTMLAVLK(2+), 18
MLLREVTKEER(2+), 17
ATLFQLDNDVQR(2+), 17
LGDFKKTLNMPR(2+), 17
QELIDFLKNSGR(2+), 16
MNEAVTIRGKER(2+), 14
ERFQLSLRQDK(2+), 14
TLAMVIETEEKR(2+), 14
m/z 894.286, 2+ 121: Scans: 2138-2162
Grouped Scans
MS/MS 2138, m/z 894.732
MS/MS 2141, m/z 894.732
MS/MS 2144, m/z 894.732
MS/MS 2147, m/z 894.732
MS/MS 2150, m/z 894.732

The *de novo* solutions are added to the peak lists tree, and you can browse down, looking for cases where the database search failed and de novo has a high score.

This looks like a promising case. But, how do we resolve the ambiguity? One of the most powerful checks is to run an error tolerant sequence tag search, and see whether the match is a modified sequence from a known protein, as we'll discuss this in the next talk.


Mascot Distiller
File
Edit
View
Processing
Analysis
Tools
Windows
Help
New Project
New Multi File Project...
Manage Multi File Project...
Open Project... Ctrl+O
Close Project
Close All
Save Project Ctrl+S
Save Project As...
Save Peak List As
Export Profile Data...
Export All Denovo Solutions...
Print... Ctrl+P
Print Preview
1 D:\...\qtof20953.raw.rov
2 D:\...\qtof20953.raw
3 D:\...\qtof10348.raw.rov
4 CL_RB192_with-compound_MSMS-A8-full-centroid.rov
5 CL_RB192_with-compound_MSMS-A8-full-centroid.mzML
6 CL_RB192_control_MSMS_Full-Centroid.rov
7 CL_RB192_control_MSMS_Full-Centroid.mzML
8 D:\...\qtof10348.raw
9 54d1c2e7be016c68__20121221_BoSo_Ecoli_SuperSilac_EtOH_Rep2_IG_HCD_LysC_2hours_03.raw.-1.rov
Exit
New Multi File project
Add File(s)...
Remove
Add Project...
Browse...
...
OK
Cancel

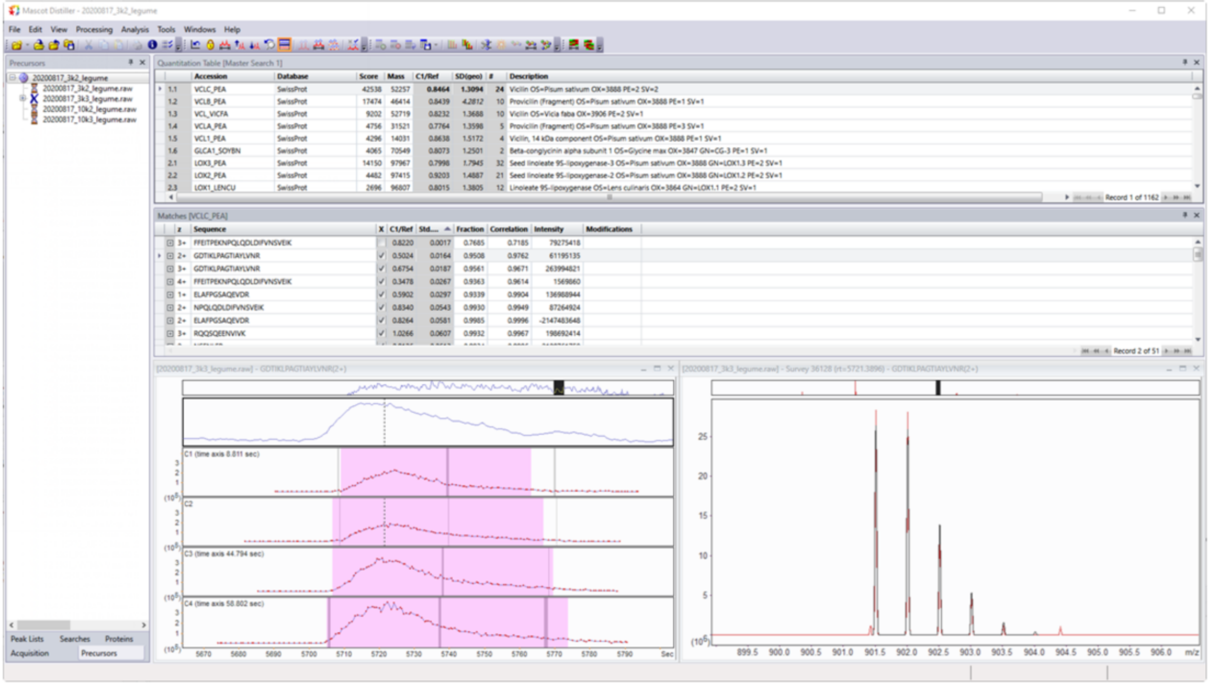
	Sample name	#	Status	Filename	Path
1		1	New	20200817_3k2_legume.raw	G:\MSData\Label free
2		1	New	20200817_3k3_legume.raw	G:\MSData\Label free
3		1	New	20200817_10k2_legume.raw	G:\MSData\Label free
4		1	New	20200817_10k3_legume.raw	G:\MSData\Label free

Distiller Project File: G:\MSData\Label free\20200817_3k2_legume.rov

Processing options: default.ThermoXcalibur.opt

Import settings and processing options from:

A Distiller project can contain more than one raw file. Choose New multi-file project from the File menu to invoke a file selection box. All the data to be processed as separate files and the search results combined.



Quantitation Table [Master Search 1]

	Accession	Database	Score	Mass	C1/Ref	SD(geo)	#	Description
1.1	VCLC_PEA	SwissProt	42538	52257	0.8464	1.3094	24	Vicilin OS=Pisum sativum OX=3888 PE=2 SV=2
1.2	VCLB_PEA	SwissProt	17474	46414	0.8439	4.2812	10	Provicilin (Fragment) OS=Pisum sativum OX=3888 PE=1 SV=1
1.3	VCL_VICFA	SwissProt	9202	52719	0.8232	1.3688	10	Vicilin OS=Vicia faba OX=3906 PE=2 SV=1
1.4	VCLA_PEA	SwissProt	4756	31521	0.7764	1.3598	5	Provicilin (Fragment) OS=Pisum sativum OX=3888 PE=3 SV=1
1.5	VCL1_PEA	SwissProt	4296	14031	0.8638	1.5172	4	Vicilin, 14 kDa component OS=Pisum sativum OX=3888 PE=1 SV=1
1.6	GLCA1_SOYBN	SwissProt	4065	70549	0.8073	1.2501	2	Beta-conglycinin alpha subunit 1 OS=Glycine max OX=3847 GN=CG-3 PE=1 SV=1
2.1	LOX3_PEA	SwissProt	14150	97967	0.7998	1.7945	32	Seed linoleate 9S-lipoxygenase-3 OS=Pisum sativum OX=3888 GN=LOX1.3 PE=2 SV=1
2.2	LOX2_PEA	SwissProt	4482	97415	0.9203	1.4887	21	Seed linoleate 9S-lipoxygenase-2 OS=Pisum sativum OX=3888 GN=LOX1.2 PE=2 SV=1
2.3	LOX1_LENCU	SwissProt	2696	96807	0.8015	1.3805	12	Linoleate 9S-lipoxygenase OS=Lens culinaris OX=3864 GN=LOX1.1 PE=2 SV=1

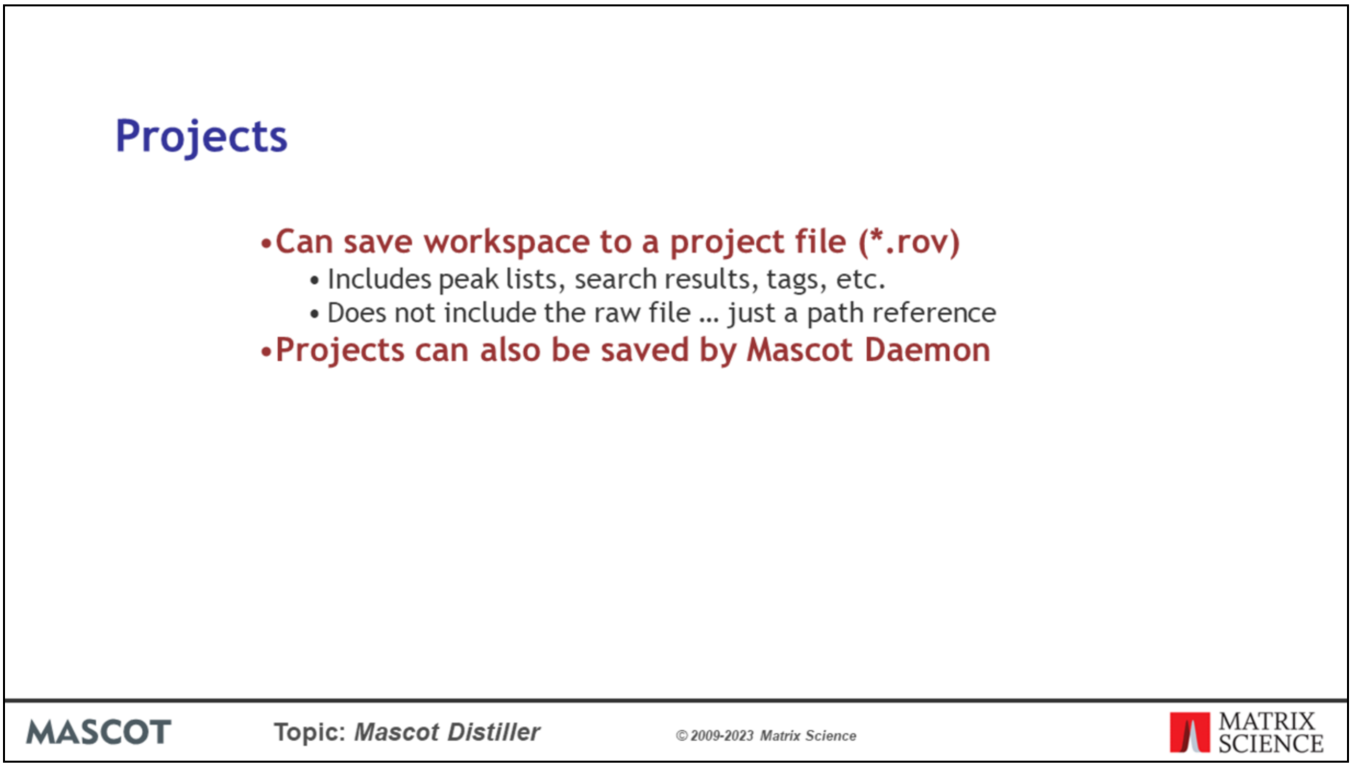
Record 1 of 1162

Matches [VCLC_PEA]

z	Sequence	X	C1/Ref	Std....	Fraction	Correlation	Intensity	Modifications
3+	FFEITPEKNPQLQDLDIFVNSVEIK	☐	0.8220	0.0017	0.7685	0.7185	79275418	
2+	GDTIKLPAGTIAYLVNR	✓	0.5024	0.0164	0.9508	0.9762	61195135	
3+	GDTIKLPAGTIAYLVNR	✓	0.6754	0.0187	0.9561	0.9671	263994821	
4+	FFEITPEKNPQLQDLDIFVNSVEIK	✓	0.3478	0.0267	0.9363	0.9614	1569860	
1+	ELAFPGSAQEVDR	✓	0.5902	0.0297	0.9339	0.9904	136988944	
2+	NPQLQDLDIFVNSVEIK	✓	0.8340	0.0543	0.9930	0.9949	87264924	
2+	ELAFPGSAQEVDR	✓	0.8264	0.0581	0.9985	0.9996	-2147483648	
3+	RQQSQEENVIVK	✓	1.0266	0.0607	0.9932	0.9967	198692414	

Record 2 of 51

Multi-file project are mainly used for label-free quantitation, which will be discussed in a separate presentation, but can also be useful for MudPIT fractions. The raw data files are processed individually, and a ‘memory efficient’ multifile project created, which reduces the memory overhead. This can then be used for quantitation.



Projects

- **Can save workspace to a project file (*.rov)**
  - Includes peak lists, search results, tags, etc.
  - Does not include the raw file … just a path reference
- **Projects can also be saved by Mascot Daemon**

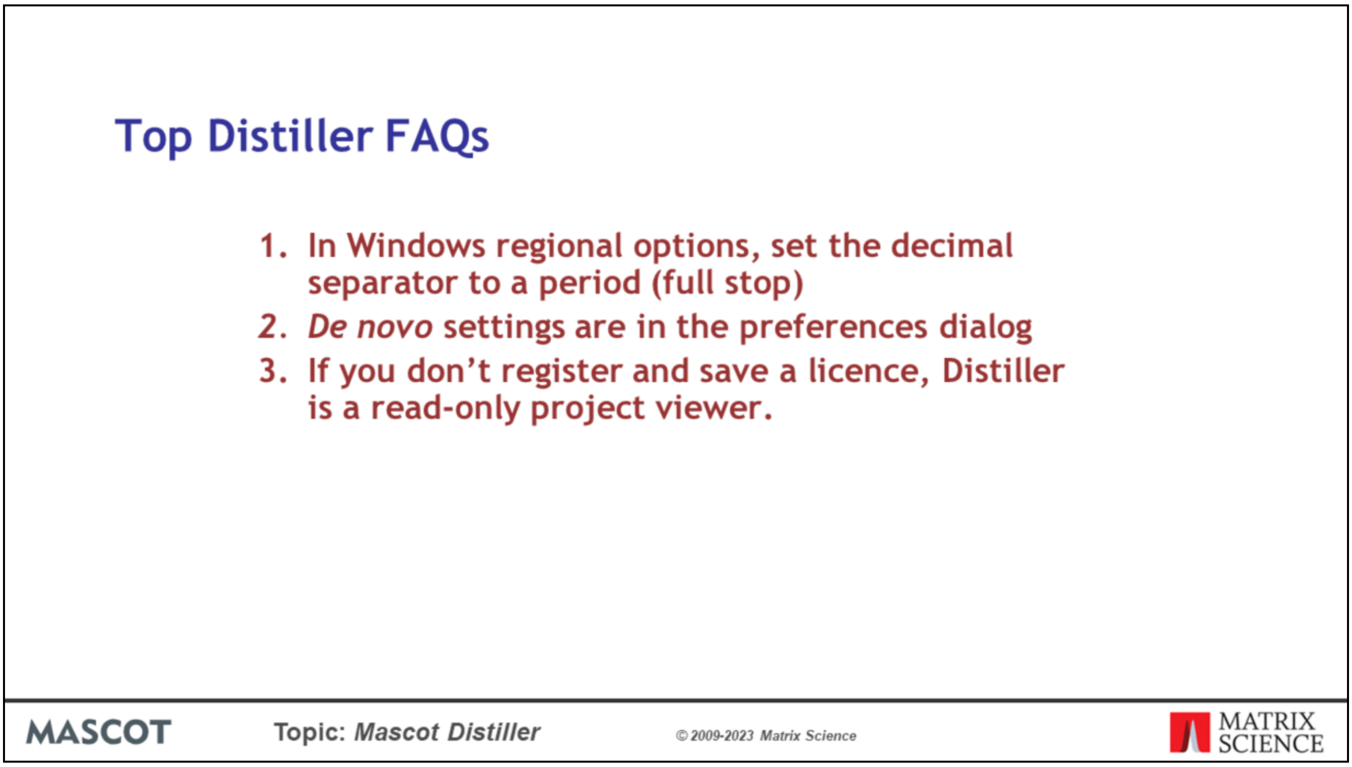
MASCOT   Topic: *Mascot Distiller*

© 2009-2023 Matrix Science

MATRIX SCIENCE

Having processed the data and maybe performed some Mascot searches and *de novo* sequencing, you will often want to be able to save everything in the workspace. You can do exactly that. The only thing that doesn’t go into the project file is the raw file itself, for space reasons. If the reference to the raw file is broken, you can easily re-attach the raw file when opening a project.

You can also choose to have Mascot Daemon save Distiller project files.



Top Distiller FAQs

1. In Windows regional options, set the decimal separator to a period (full stop)
2. *De novo* settings are in the preferences dialog
3. If you don’t register and save a licence, Distiller is a read-only project viewer.

MASCOT Topic: *Mascot Distiller* © 2009-2023 Matrix Science MATRIX SCIENCE

Finally, a few of the most frequent technical support questions:

1. Some of the file access libraries are not fully localised. In Windows regional options, you need to set the decimal separator to a period (full stop)

2. *De novo* settings are in the preferences dialog. Most of these settings will be the same as on the Mascot search tab

3. If you don’t register and save a licence, Distiller is a read-only project viewer