New features in Mascot Distiller - MS^E, MS3 reporter ion quant, and more

© 2016 Matrix Science

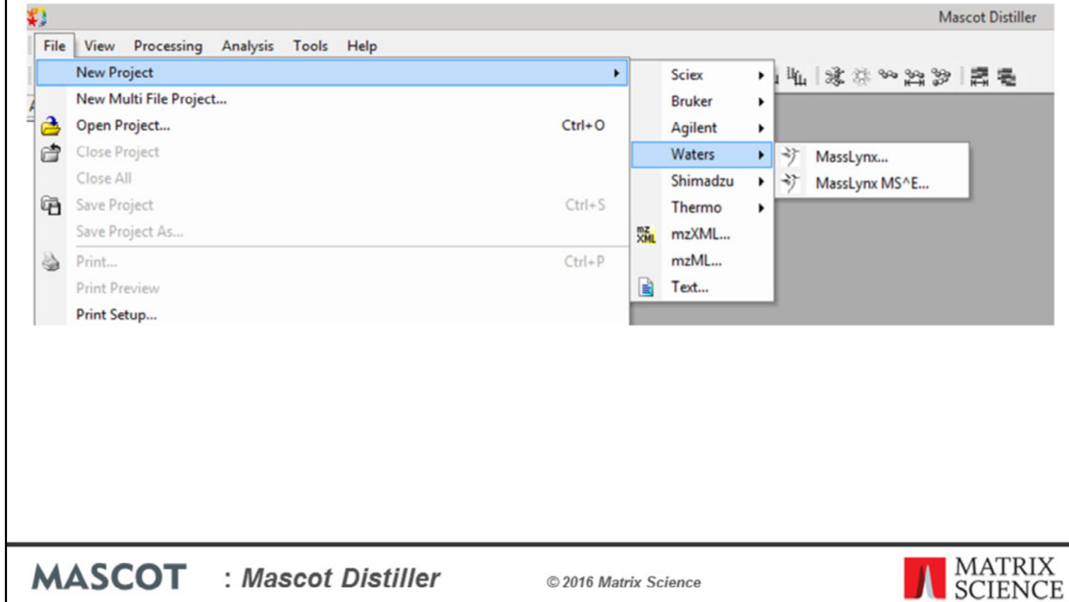
New features in Mascot Distiller 2.6

1. Waters MS^E support
2. MS3 reporter ion quantitation
3. Merging scan types
4. Sciex – vendor peak picking
5. New Peak List options
6. Quantitation 1 hit wonders

MASCOT : *Mascot Distiller* © 2016 Matrix Science MATRIX SCIENCE

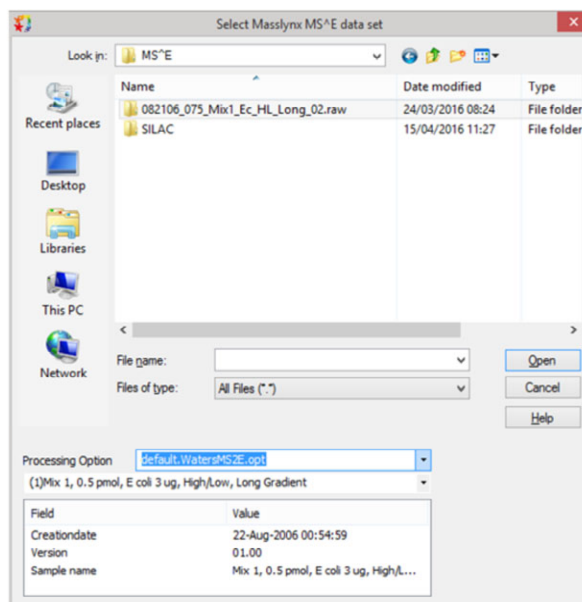
We've introduced a number of new features into the forthcoming Mascot Distiller 2.6 release. We'll start by taking a look at the newly added support for Waters MS^E datasets

Waters MS^E support



When you create a new project in Mascot Distiller 2.6 you'll see a new option under the Waters sub-menu – MassLynx MS^E. To create a new MS^E project, simply select this option.

Waters MS^E support

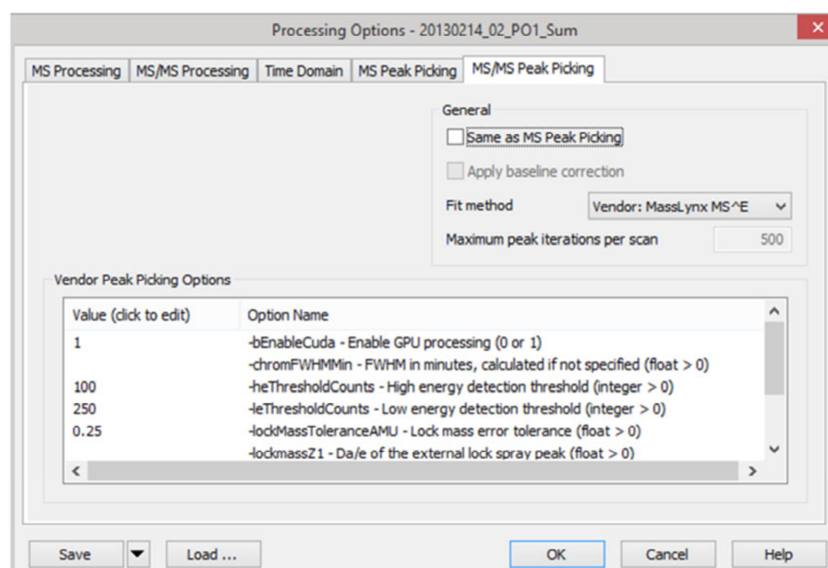


MASCOT : *Mascot Distiller* © 2016 Matrix Science

And select the MS^E raw data folder in the file dialog.

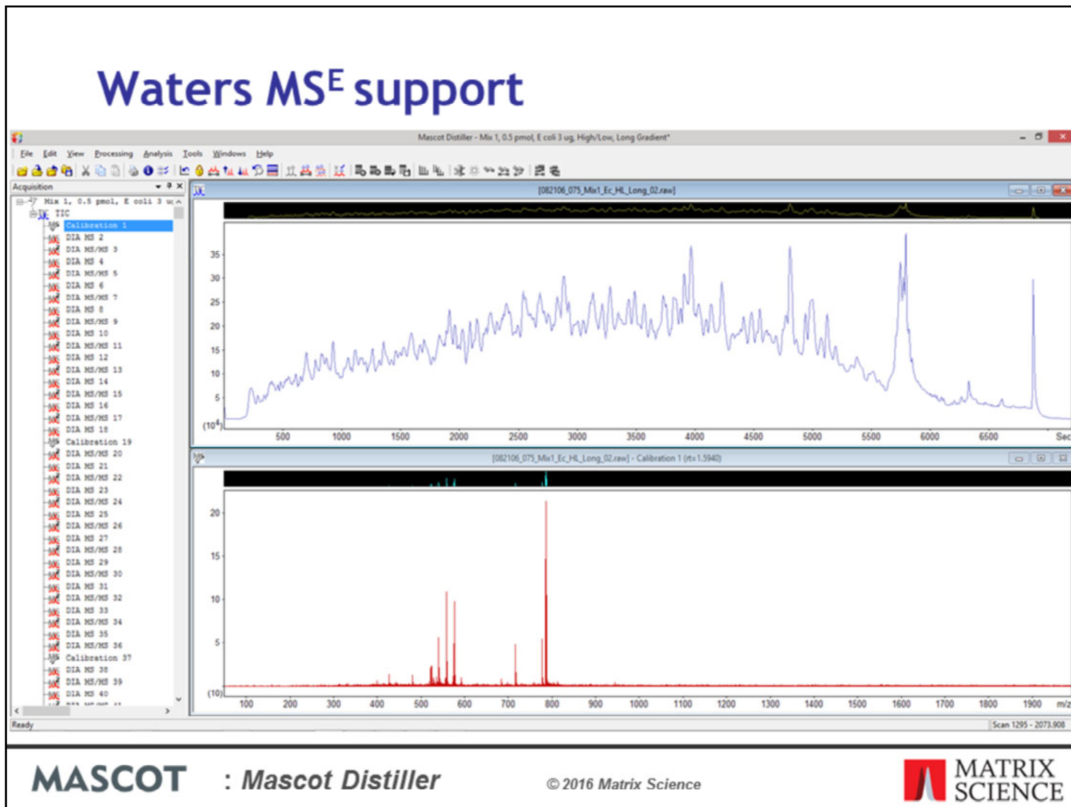
This allows you to open and process MassLynx MS^E datasets within Mascot Distiller – including using them for precursor and label free quantitation experiments.

Waters MS^E support

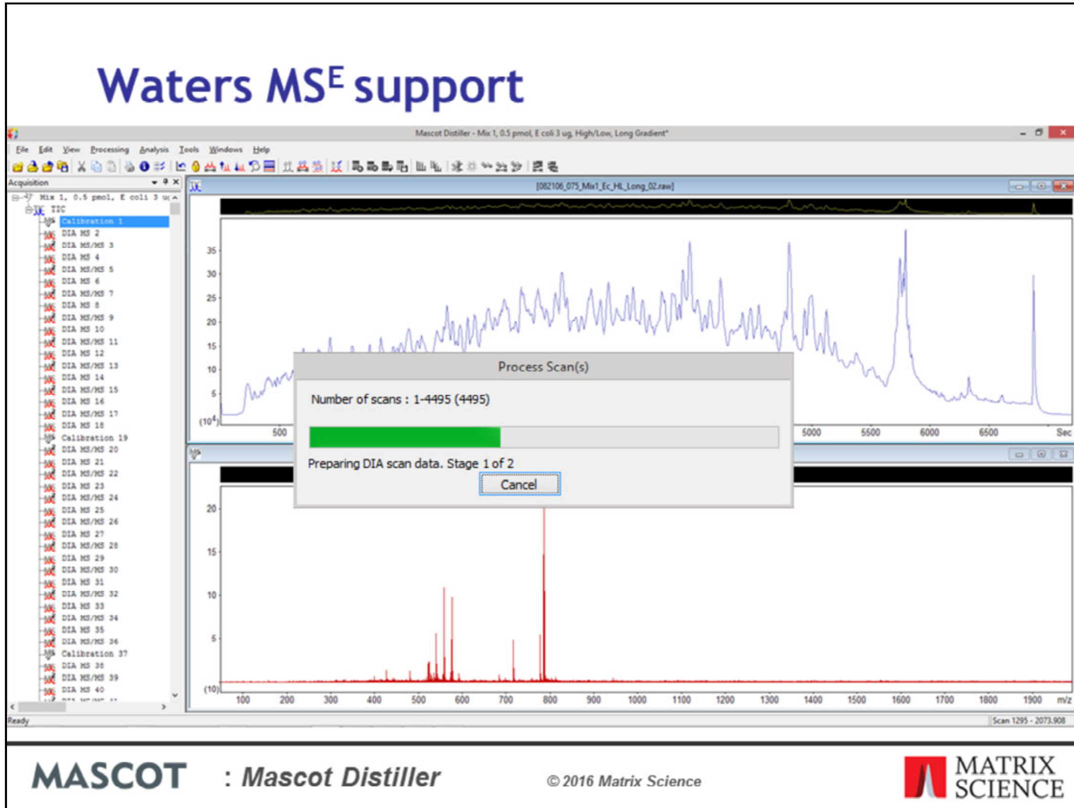


MASCOT : *Mascot Distiller* *© 2016 Matrix Science* MATRIX SCIENCE

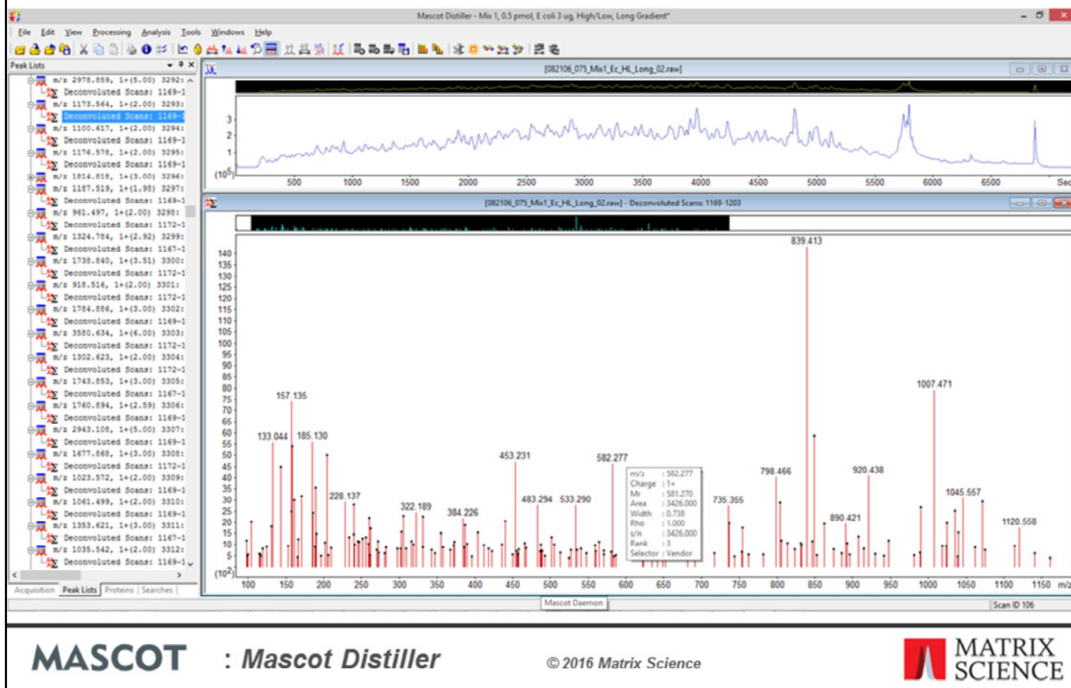
We're using the Waters libraries to access and deconvolute the MS^E peaklists, while precursor processing is done internally. So if you are setting up MS^E processing options in Mascot Distiller, on the MS/MS Peak Picking settings tab you'll see a new drop down menu for 'Fit method'. This allows you to select the Vendor:MassLynx MS^E fit method instead of the normal Isotope Distribution option.



When we open an MS^E dataset, this is what the acquisition tree looks like, with a mixture of DIA MS, MS/MS scans along with Calibration scans

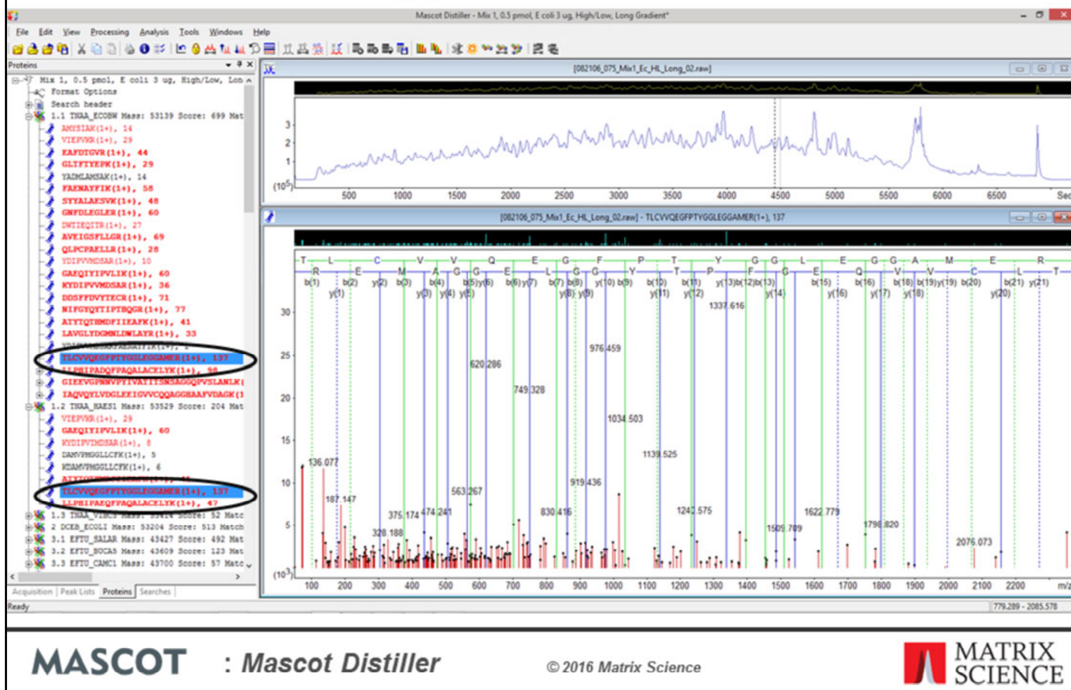


We carry out peak detection on the dataset as normal



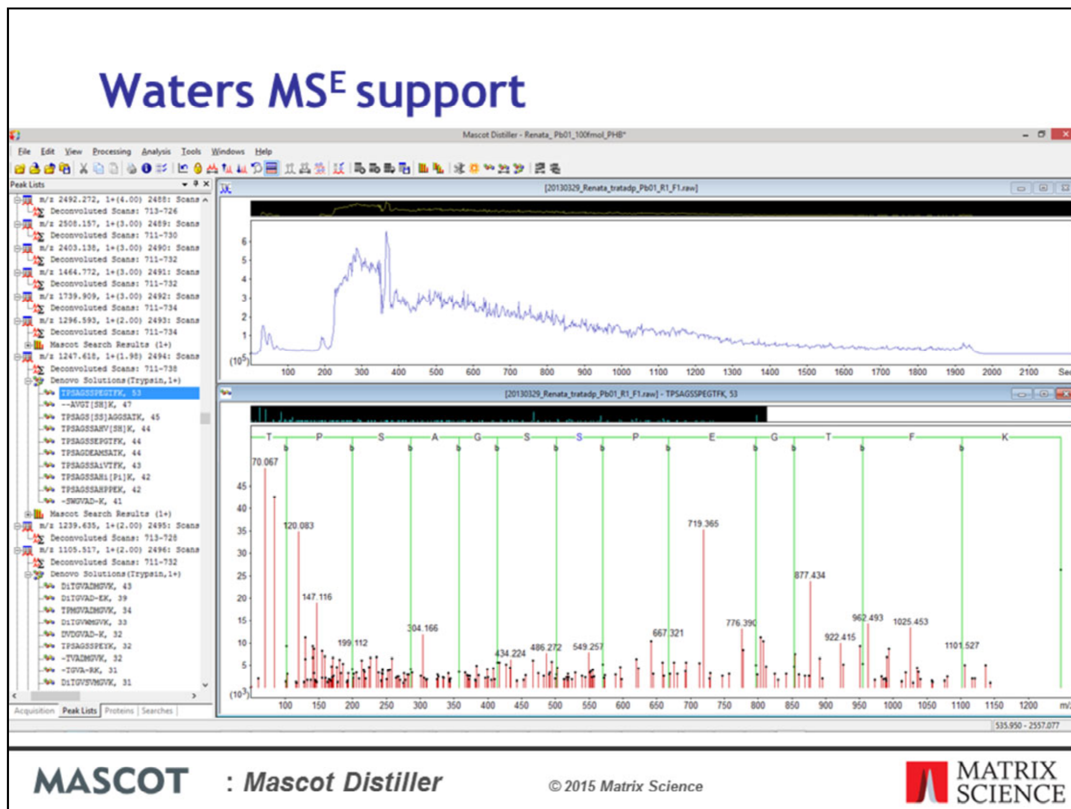
And the generated peak lists dataset looks like this, with our deconvoluted scans under the precursor.

Waters MS^E support



MASCOT : *Mascot Distiller* © 2016 Matrix Science

We can now submit the peaklists to Mascot for a database search just like any other peaklist. One additional minor improvement you can see in this slide is the background highlighting of peptides. When we select a peptide match, such as this one here, it is hightlighted in the Proteins tree in any proteins it matches into.



In addition to this, we can use the extra functionality provided by Mascot Distiller with the processed MS^E data – for example, the Mascot Distiller Search Toolbox provides a powerful de novo sequencing tool. This dataset is from a species of fungi which is poorly represented in the sequence databases - Here we have an example where the de novo tool has found a good match on one of the MS^E spectra where the standard database search found no significant matches. A quick BLAST of the denovo sequence result found a match to a sequence from a related species.

Waters MS^E support

MS1 based Quantitation support:

- SILAC
- 18O
- Metabolic 15N
- Dimethylation etc

MASCOT : *Mascot Distiller* © 2016 Matrix Science

One of the biggest advantages of processing the MS^E data in Mascot Distiller is that you can now use these datasets for the precursor and label free quantitation protocols supported by the Mascot Distiller quantitation toolbox. Here we have an example of a SILAC MS^E quantitation dataset, processed searched and quantified with Mascot Distiller and Mascot Server.

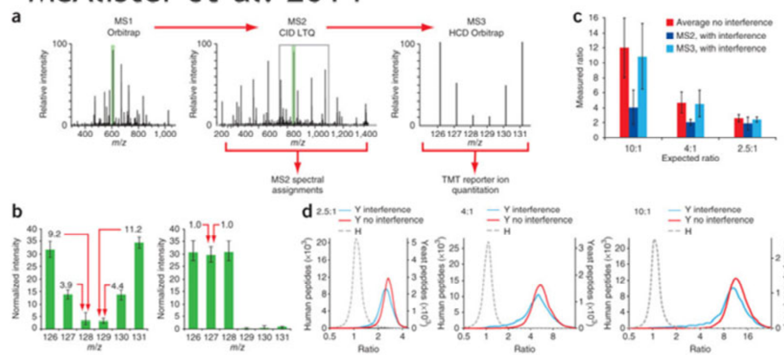
New features in Mascot Distiller 2.6

1. Waters MS^E support
2. MS3 reporter ion quantitation
3. Merging scan types
4. Sciex – vendor peak picking
5. New Peak List options
6. Quantitation 1 hit wonders

© 2016 Matrix Science

MS3 reporter ion quantitation

- Strategy to reduce problems of interfering ions in reporter ion quantitation
  - Ting *et al.* 2011
  - McAlister *et al.* 2014

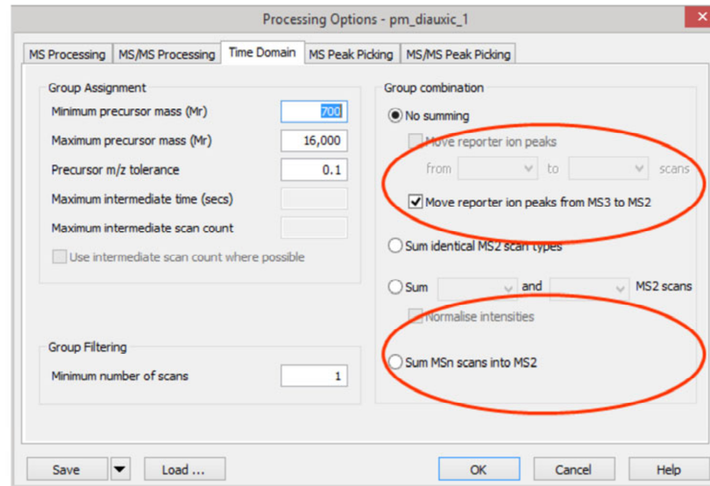


MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics
Lily Ting1, Ramin Rad1, Steven P Gygi1, Wilhelm Haas1
Nature Methods 8, 937-940 (2011)

© 2016 Matrix Science

A common problem encountered when carrying out reporter ion quantitation methods, such as TMT or iTRAQ, is that of interfering ion signals in the reporter region. One approach which can be taken to reducing the issue is to reisolate the most abundant ion in the MS/MS spectrum and refragment it – the resulting reporter ion signals in the MS3 signal are then almost exclusively from the target peptide.

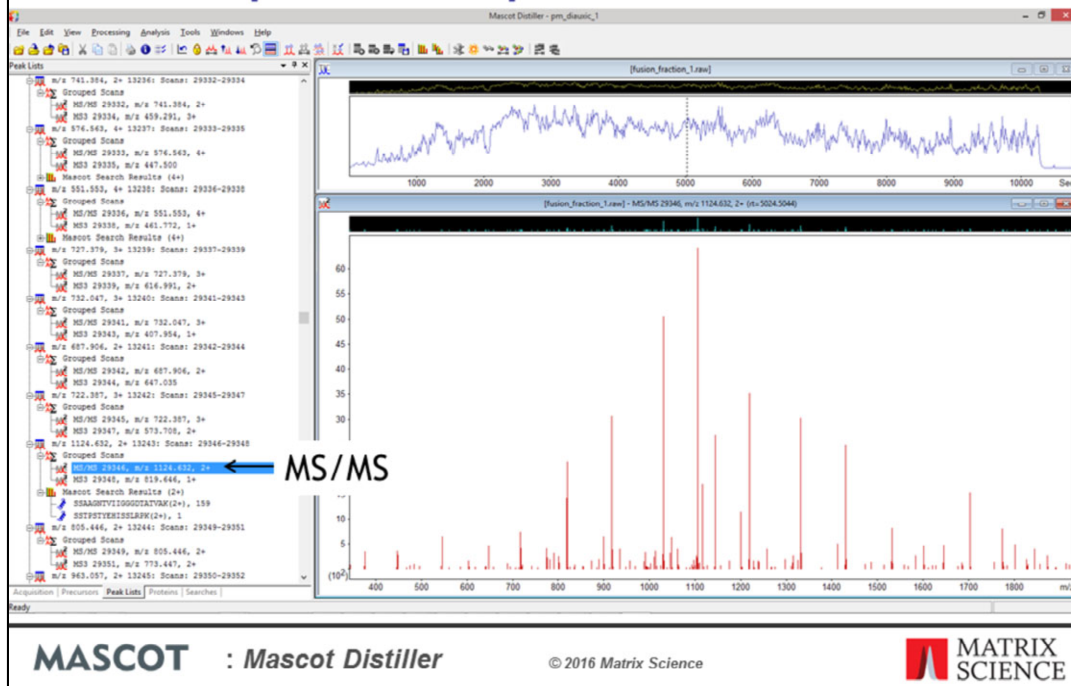
MS3 reporter ion quantitation



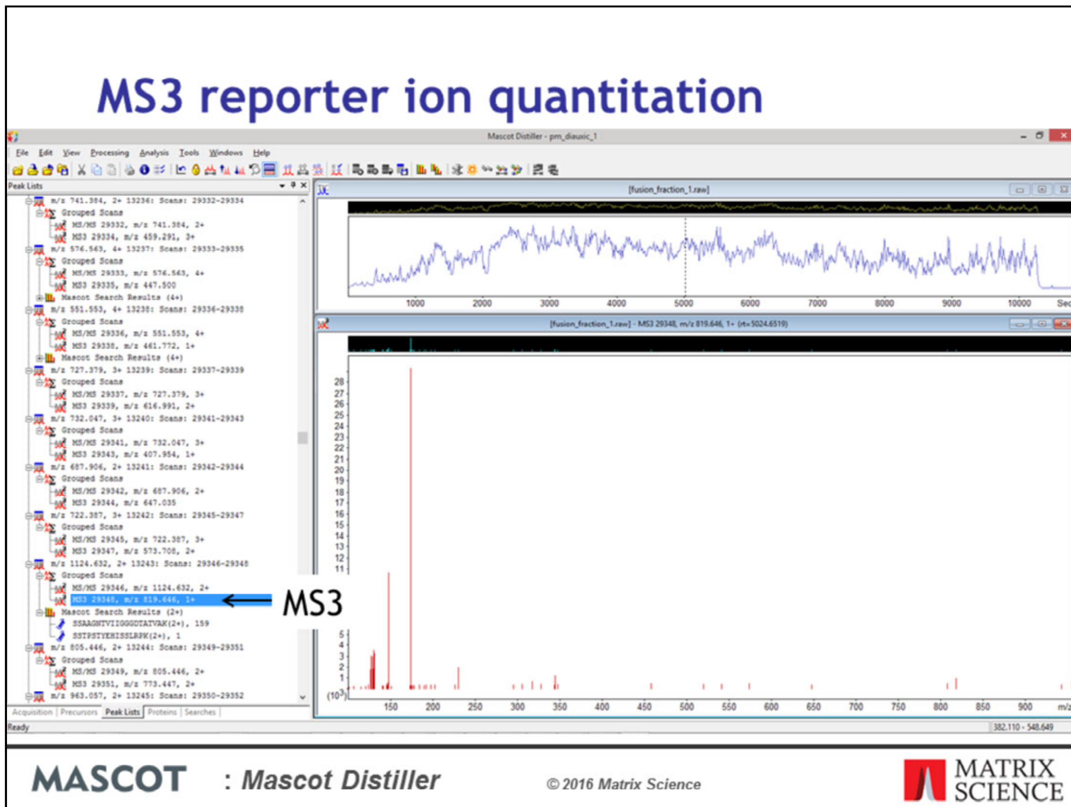
MASCOT : *Mascot Distiller*

© 2016 Matrix Science

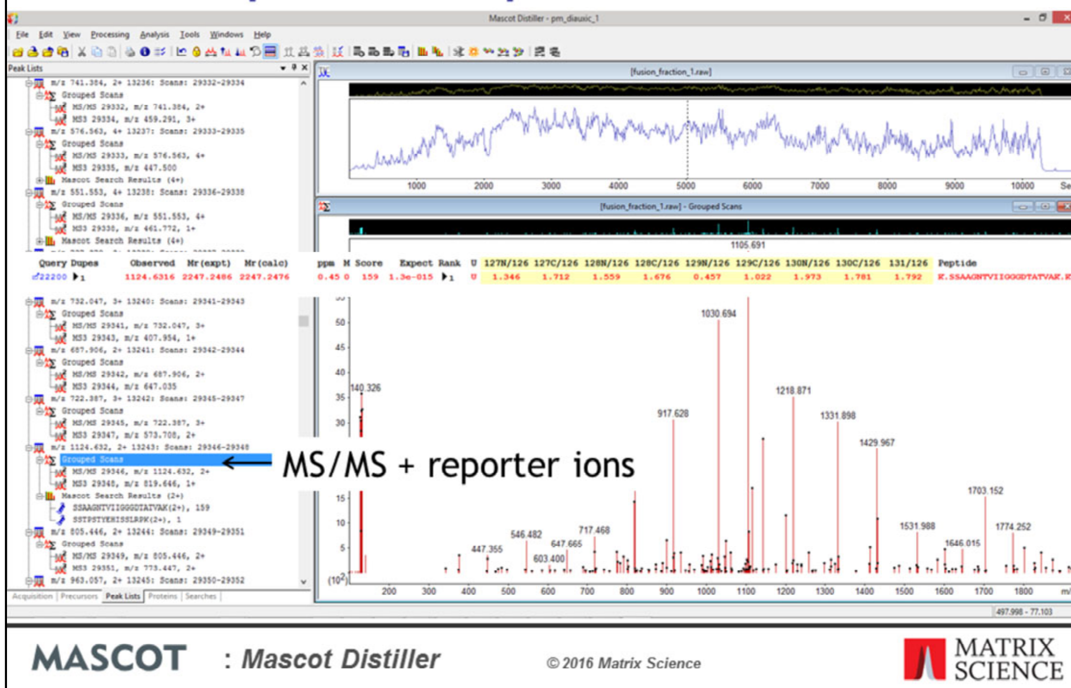
To support this, we've added new options on the Time Domain section of the Processing Options dialog in Mascot Distiller. Under the 'Group combination' section, if we have 'No summing' selected, we can choose to move the reporter ion peaks from the MS3 scans to the MS2 scans. Alternatively, you can sum the entire MS3 scan into the MS2 scan.



Once we've carried out peak detection, on the peak lists tab we have our MS2 and MS3 scans grouped together. This is the MS/MS scan – notice that it doesn't contain any reporter ions.



And this is the MS3 scan containing the reporter ions. If we click on the grouped scans node



This is the MS/MS scan with the MS3 report ion region merged into it – this is the peaklist that we searched in Mascot. As you may be able to see, this peak list got a very impressive match with a score of 159. Here is the result in Mascot with the quantitation ratios.

New features in Mascot Distiller 2.6

1. Waters MS^E support
2. MS3 reporter ion quantitation
3. **Merging scan types**
4. Sciex - vendor peak picking
5. New Peak List options
6. Quantitation 1 hit wonders

MASCOT : *Mascot Distiller*

© 2016 Matrix Science

MATRIX SCIENCE

In addition to merging MS3 scans into the parent MS/MS scans, Mascot Distiller 2.6 includes options for summing or merging scans together based on scan type.

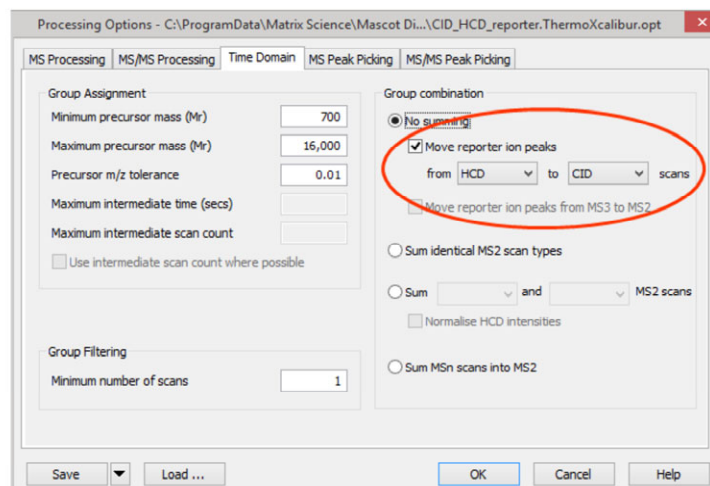
Merging different scan types

- Multiple scan types for the same precursor
  - E.g. CID and HCD scans from Thermo Orbitrap
- Often wish to search merged spectra
  - E.g. CID with reporter ions in HCD scan
    - Splicing just the reporter ion information into the CID scans

MASCOT : *Mascot Distiller* © 2016 Matrix Science MATRIX SCIENCE

Some instruments can produce raw data files with multiple different types of scan data for the same precursor in them – a common example of this would be an experiment on a Thermo Orbitrap instrument raw data file containing both CID and HCD scans for the same precursor. Often, you would often want to sum these spectra before carrying out a database search – For example, a common experiment to improve the quality of report ions when using quantitation methods such as iTRAQ or TMT involves carrying out DDA of CID scans on the peptide precursor, and then further fragmenting the same precursor using HCD to get spectral information in the reporter ion region. In this case you'd want to merge the information from the reporter ion information from the HCD scan into the CID scan, and then search the merged peaklist.

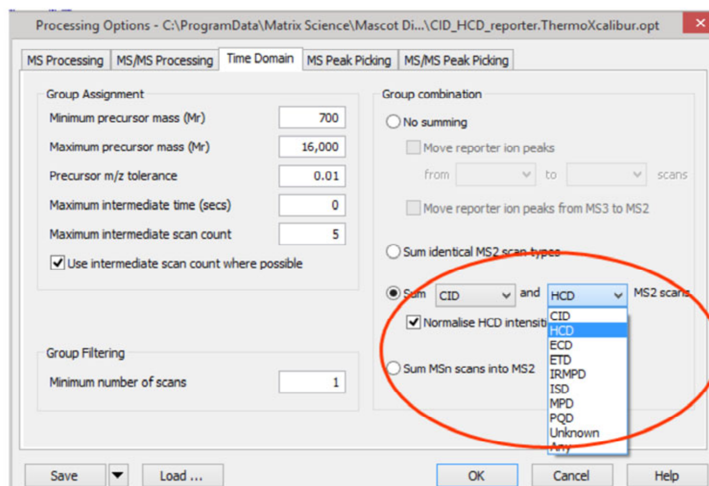
Merging different scan types



© 2016 Matrix Science

In Mascot Distiller 2.6 we've introduced the ability to merge different scan types. The options to do this can be found on the Time-Domain tab of the processing options dialog in the Group combination section. If you are interested in only moving reporter ions from one scan-type to another, you can select 'No summing' with the 'Move reporter ion peaks' checkbox checked, as shown here. These processing options would move the reporter ion region specified under the 'single peak picking' region of the MS/MS Peak Picking options from the HCD into the CID scans.

Merging different scan types

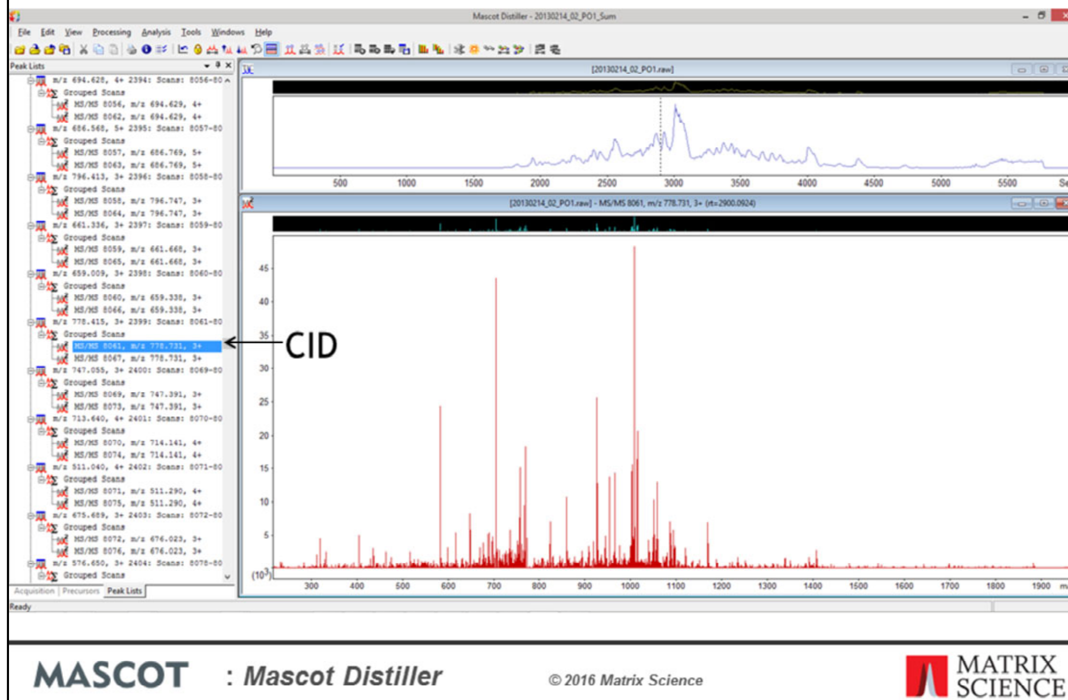


MASCOT : *Mascot Distiller* © 2016 Matrix Science MATRIX SCIENCE

Previously this could merge complete scans together on the basis of precursor m/z and retention time – now you can also use the scan type.

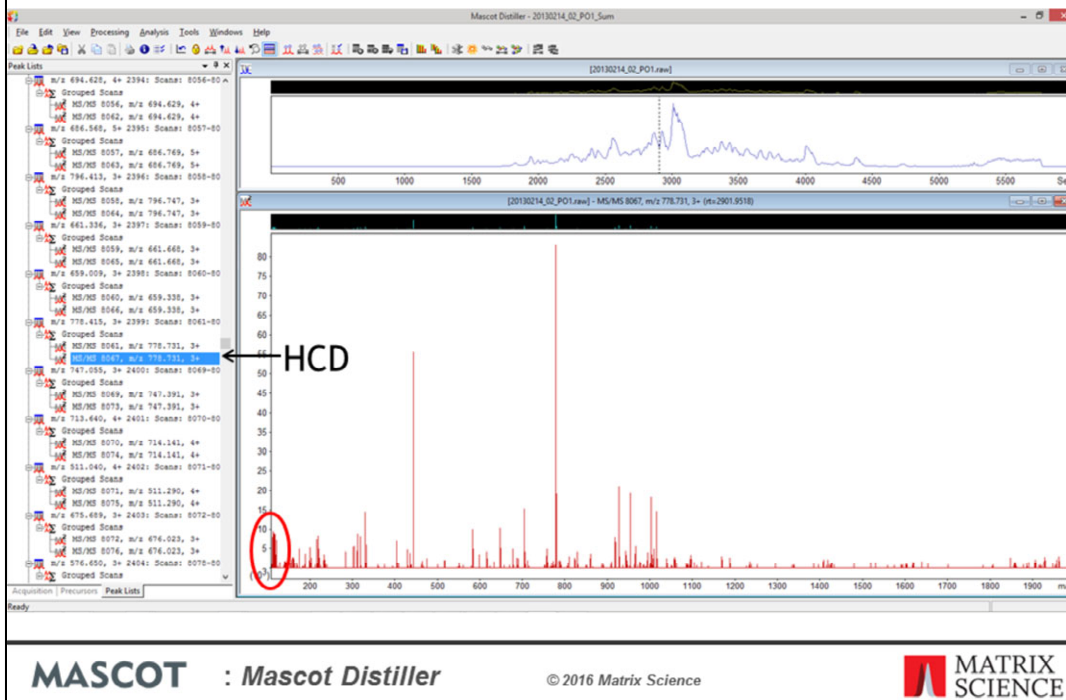
So you can also sum complete spectra together of any or of specified types, with options for normalising the intensities of the scan type being summed in to the ‘parent’ scan.

Merging different scan types

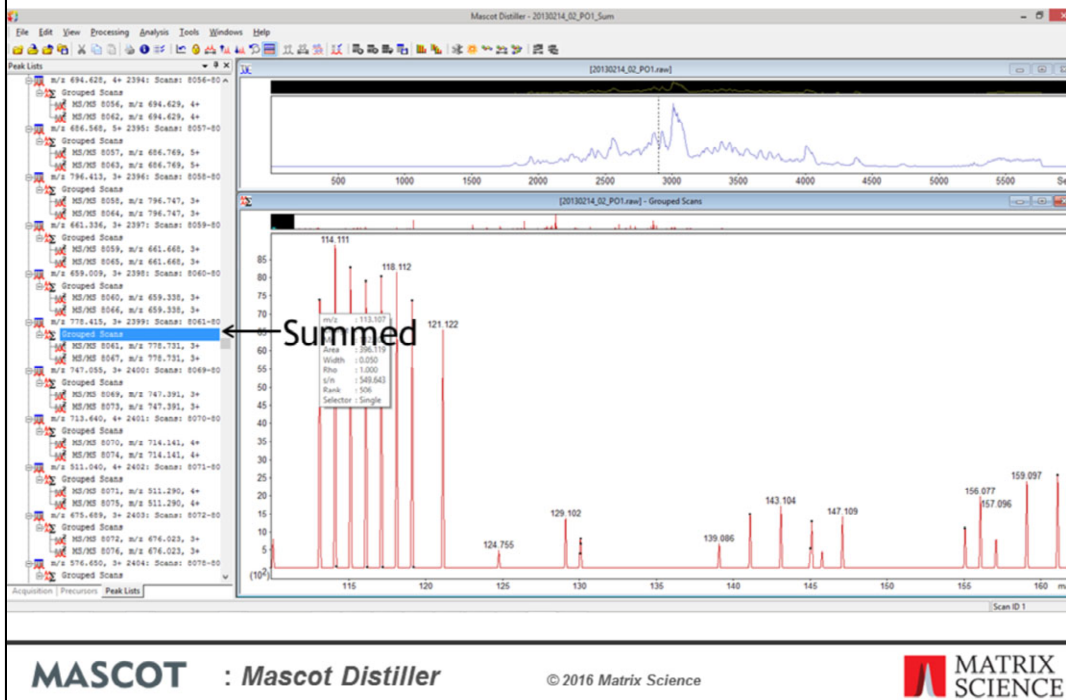


Data taken from http://www.ebi.ac.uk/pride/archive/projects/PXD000503

Once we've carried out peak detection this is what the Peak Lists tree looks like in Distiller, with a Grouped Scans node, under which we have the source scans. This is the CID scan. This is an iTRAQ 8-Plex dataset, and notice that we don't have any peaks in the spectrum below an m/z ratio of about 200. So no reporter ions.



This is the HCD scan. This is where the iTRAQ 8-Plex reporter ions are.



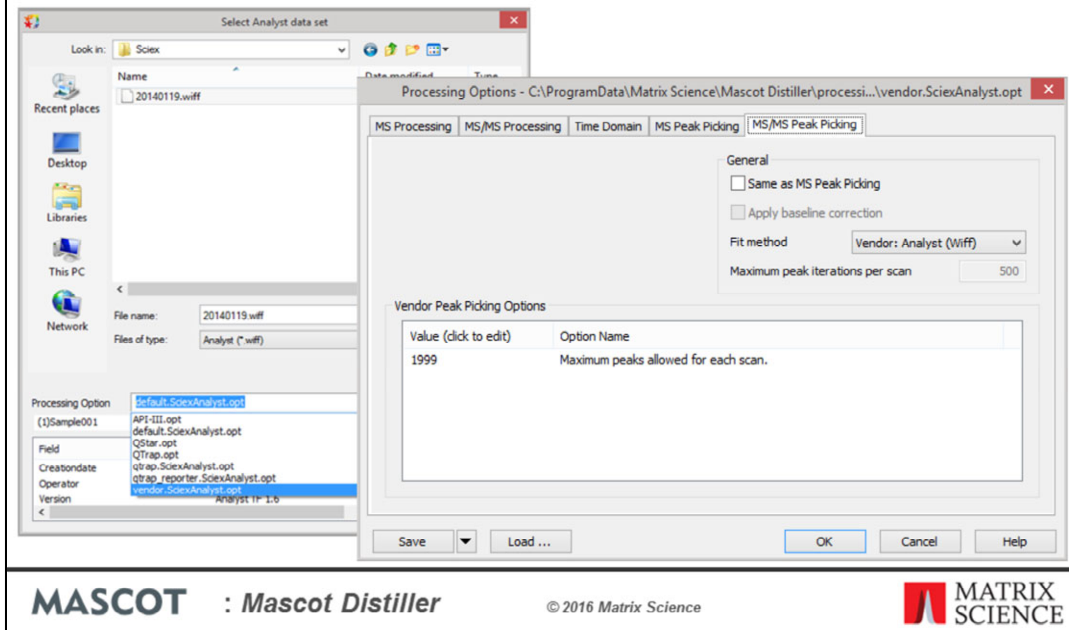
And this is the Grouped Scan with the picked peaks – This contains information from both the HCD and CID scans. I've zoomed in on the report ion region here.

New features in Mascot Distiller 2.6

1. Waters MS^E
2. MS3 reporter ion quantitation
3. Merging scan types
4. Sciex – vendor peak picking
5. New Peak List options
6. Quantitation 1 hit wonders

© 2016 Matrix Science

Sciex: Vendor MS/MS peak picking



When you create a new project for Sciex .wiff datafiles, you'll see that we've added an additional standard processing options file – vendor. If you select this option, then MS peak detection is carried out by Mascot Distiller, while MS/MS peak picking uses the vendor peak detection. You can select this from the MS/MS peak picking tab on the Processing Options dialog window in Mascot Distiller, so you can switch to using the vendor peak detection for your own, tweaked, processing options.

Sciex: Vendor MS/MS peak picking

- Advantage - Speed
  - Vendor peak picking has already been carried out
- Disadvantage - can’t tweak MS/MS parameters
  - Vendor peak picking has already been carried out
- Precursor peak picking by Mascot Distiller
  - MS/MS peak picking by Mascot Distiller
  - MS/MS peak picking from vendor
- Mascot Distiller: ~22 minutes
- Vendor: 63 seconds

MASCOT : *Mascot Distiller* © 2016 Matrix Science

The big advantage of this is speed – because the vendor peak detection has already been carried out, we’re just pulling the values out of the .wiff file. The disadvantage of this is that you cannot of course further tweak the MS/MS peak detection parameters within Mascot Distiller.

As a comparison, I carried out MS/MS peak picking using either Mascot Distiller or the vendor peak picking on the same raw datafile. In both cases, precursor peak detection was carried out by Mascot Distiller. As you can see, using the vendor MS/MS peak detection was considerably faster.

Sciex: Vendor MS/MS peak picking

MASCOT : *Mascot Distiller*

© 2016 Matrix Science

The test dataset I used was a SILAC experiment. Because precursor peak picking and processing was carried out by Mascot Distiller, even though vendor peak picking was used for the MS/MS spectra, we can still carry out SILAC quantitation using Mascot Distiller. So, in the screenshot shown here, we've used Mascot Distiller for precursor peak detection, then used vendor peak detection for the MS/MS spectra. Searched the peaklists in Mascot and carried out SILAC quantitation in Mascot Distiller.

New features in Mascot Distiller 2.6

1. Waters MS^E support
2. MS3 reporter ion quantitation
3. Merging scan types
4. Sciex – vendor peak picking
5. **New Peak List options**
6. Quantitation 1 hit wonders

© 2016 Matrix Science

New Peak List options

Instrument Definition	Peptide matches (1% FDR)
Composite EDT+CID instrument	7740
Separate instrument definitions	8084

MASCOT : *Mascot Distiller* © 2016 Matrix Science MATRIX SCIENCE

If your raw data files contains different scan types (e.g. CID and ETD scans), then you might want to search the generated peaklists using different ions series depending on the source scan type. Normally, to search CID and ETD scans together, we recommend using a composite CID+ETD instrument definition which contains the ions series for both scan types.

However, the Mascot Generic peaklist format has supports setting the instrument type at the level of each individual MS/MS peaklist. In Mascot Distiller 2.6, on the ‘Peak List Format’ tab under preferences, you’ll find an additional option which allows you to automatically set the INSTRUMENT parameter in the exported peaklists of a selected scan type. For these spectra, when the peaklist is generated, the INSTRUMENT parameter value will override the option selected on the Mascot search form.

Searching the different scan types with the correct instrument definitions can give you better sensitivity – for example, searching this dataset with separate EDT and CID instruments yields a 4.4% increase in the number of significant peptide matches at 1% FDR when compared to searching with the composite instrument definition.

New Peak List options

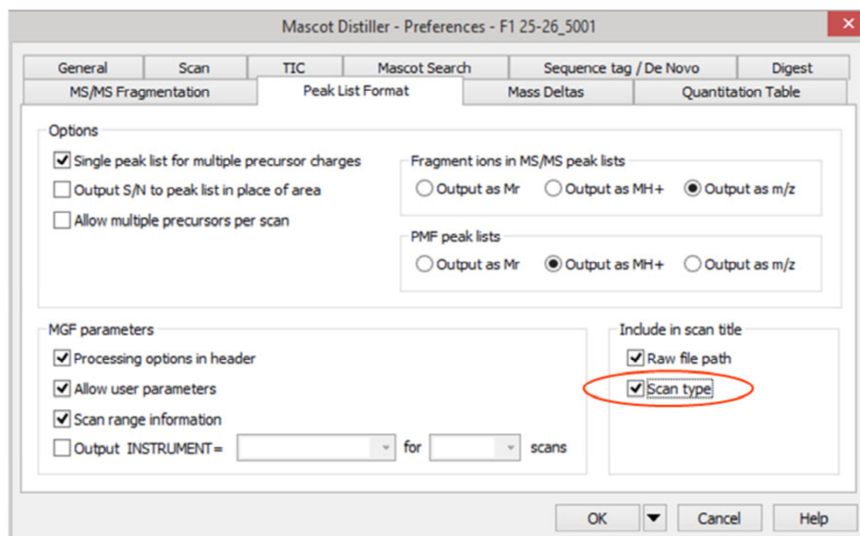
- Structured file information in PSI:CV terms output to MGF header
  - _DISTILLER_SOURCE_TYPE
  - _DISTILLER_INSTRUMENT_MODEL
- Optionally, Scan Type PSI:CV terms can be output to individual scan titles

MASCOT : *Mascot Distiller* © 2016 Matrix Science MATRIX SCIENCE

In Mascot 2.6, if you select the MGF or mzData peak list formats, additional structured information about the peak list can be output using the controlled vocabulary defined by the Proteomics Standards Initiative. In the header of an mgf file generated by Distiller 2.6, you'll find these two new labels, _DISTILLER_SOURCE_TYPE and _DISTILLER_INSTRUMENT_MODEL, which will give the source and instrument model as the PSI:CV term.

In addition, you can include scan type as PSI_CV terms in the individual scan titles

New Peak List options



If you want to include the scan type in the scan titles, go to the ‘Peak List Format’ tab on the ‘Preferences’ dialog and ensure that the new ‘Scan type’ checkbox is checked before you export the peaklist.

New Peak List options

```
_DISTILLER_MDRO_VERSION=2.5.0.243
_DISTILLER_ORIGINAL_MDRO_VERSION=2.5.0.0
_DISTILLER_RAWFILE={1}D:\MSData\Example datasets\Chalkley_MCP_2005_4_1189_1193\wiff\F1 5uLUCSF.wiff
_DISTILLER_SOURCE_TYPE={1}MS:1000073
_DISTILLER_INSTRUMENT_MODEL={1}MS:1000656
_DISTILLER_SINGLE_PRECURSOR_PER_PEAKLIST=1
_DISTILLER_USE_SNRATIO_AS_INTENSITY=0
_DISTILLER_FRAGMENT_IONS_MASS_TYPE=m/z
_DISTILLER_PEPTIDE_IONS_MASS_TYPE=mh+

BEGIN IONS
TITLE=1: Scan 386 (rt=101.368, p=0, c=96, e=1) [D:\MSData\Example datasets\Chalkley_MCP_2005_4_1189_1193\wiff\F1 5uLUCSF.wiff] MS:1000133 MS:1000084
PEPMASS=453.34201 1241
CHARGE=2+
SCANS=386
RAWSCANS=pd0cy96ex1
RTINSECONDS=101.368
89.03587    0.014532215
97.031439   0.0069116147
102.06378   0.0067988051
107.04769   0.006977926
114.07452   0.48558099
117.0776    0.0077467478
119.01481   0.0077416586
123.09996   0.058020235
124.12087   0.023786212
```

MASCOT : *Mascot Distiller*

© 2016 Matrix Science

MATRIX SCIENCE

Here we have an example from a peaklist file with the new options included. Here we have the source type and instrument model in PSI:CV terms

And appended to the end of the scan title we have two terms for the scan type, also in PSI:CV

New features in Mascot Distiller 2.6

1. Waters MS^E support
2. MS3 reporter ion quantitation
3. Merging scan types
4. Sciex – vendor peak picking
5. New Peak List options
6. Quantitation 1 hit wonders

© 2016 Matrix Science

Quantitation 1 hit wonders

- In Mascot Distiller 2.5 or earlier
  - Count of number of quantified matches to sequence
  - Identical peptide sequences counted multiple times
- In Mascot Distiller 2.6
  - Count of distinct sequences quantified

MASCOT : *Mascot Distiller* © 2016 Matrix Science MATRIX SCIENCE

We've made some changes to one hit wonder handling for protein quantitation in Mascot Distiller 2.6. In Mascot Distiller 2.5 or earlier, the number of quantified peptides reported in the protein quantitation information was the total number of quantified matches to the protein sequence, with identical peptide sequences being counted multiple times – so if you had matches to the same peptide sequence at different precursor charge states, that sequence would be counted multiple times. In Mascot Distiller 2.6, we've changed this so that it is now the count of distinct peptide sequences matching to the protein which is reported.

Quantitation 1 hit wonders

Quantitation Table

	Accession	Score	Mass	M/L	SD(geo)	#	H/L	SD(geo)	#	Description
876	NMT1_HUMAN	187	57112				0.8725	1.2398	2	Glycylpeptide N-tetradecanoyltransferase 1 OS=Homo sapiens GN=NMT1 PE=1 ...
877	SSRA_HUMAN	187	32216	1.4310	1.1677	2	2.0402	1.0054	2	Translocon-associated protein subunit alpha OS=Homo sapiens GN=SSR1 PE=1 ...
878	DNJA2_HUMAN	187	46344	1.2524	1.0273	2	1.4393	1.1691	3	DnaJ homolog subfamily A member 2 OS=Homo sapiens GN=DNAJA2 PE=1 SV=1
879	WDFY1_HUM...	187	47320	1.1302	1.0744	2	1.5029	1.1698	2	WD repeat and FYVE domain-containing protein 1 OS=Homo sapiens GN=WDFY...
880	NPTN_HUMAN	187	44702	0.9985	1.1609	4	**1.2151**	**1.0336**	3	Neuroplastin OS=Homo sapiens GN=NPTN PE=1 SV=2
881	SMAP_HUMAN	186	20378				0.9902	1.1187	2	Small acidic protein OS=Homo sapiens GN=SMAP PE=1 SV=1
882	BIEA_HUMAN	185	33693	**1.1570**	**1.0008**	2	1.1441	1.1321	3	Biliverdin reductase A OS=Homo sapiens GN=BLVRA PE=1 SV=2
883	LRP10_HUMAN	185	78027	0.6679	1.5087	2	1.2244	1.0847	2	Low-density lipoprotein receptor-related protein 10 OS=Homo sapiens GN=LRP1...
884	NRBP_HUMAN	185	60378	0.8153	1.1963	3	1.0048	1.0500	3	Nuclear receptor-binding protein OS=Homo sapiens GN=NRBP1 PE=1 SV=1
885	PSMD4_HUMAN	184	40940	**1.1684**	**1.0105**	4	**1.2834**	**1.1653**	4	26S proteasome non-ATPase regulatory subunit 4 OS=Homo sapiens GN=PSMD...
886	CNTP1_HUMAN	183	158220	0.9012	1.2492	2				Contactin-associated protein 1 OS=Homo sapiens GN=CNTNAP1 PE=1 SV=1
887	IDH3A_HUMAN	183	40023	1.1576	1.0703	2	1.0325	1.2972	4	Isocitrate dehydrogenase [NAD] subunit alpha, mitochondrial OS=Homo sapiens ...

Matches (877 SSRA_HUMAN Mass: 32215 Score: 187 Matches: 6)

		z	Sequence	X	M/L	Std.Err	X	H/L	Std.Err	Fraction	Correlation	Intensity	Modification
1	+	2+	GTEDFIVESLDASFR	☑	1.5968	0.10719	☑	2.0480	0.06616	0.1928	0.9696	7029959	
2	+	3+	GTEDFIVESLDASFR	☑	1.2824	0.19464	☑	2.0323	0.15468	0.3200	0.9910	528181	

Mascot Distiller 2.5

Quantitation Table

	Accession	Score	Mass	M/L	SD(geo)	#	H/L	SD(geo)	#	Description
876	NMT1_HUMAN	187	57112				0.8725	1.2398	2	Glycylpeptide N-tetradecanoyltransferase 1 OS=Homo sapiens GN=NMT1 PE=1 ...
877	SSRA_HUMAN	187	32216							Translocon-associated protein subunit alpha OS=Homo sapiens GN=SSR1 PE=1 ...
878	DNJA2_HUMAN	187	46344	1.2524	1.0273	2	1.4393	1.1691	3	DnaJ homolog subfamily A member 2 OS=Homo sapiens GN=DNAJA2 PE=1 SV=1
879	WDFY1_HUM...	187	47320	1.1302	1.0744	2	1.5029	1.1698	2	WD repeat and FYVE domain-containing protein 1 OS=Homo sapiens GN=WDFY...
880	NPTN_HUMAN	187	44702	0.9985	1.1609	4	**1.2151**	**1.0336**	3	Neuroplastin OS=Homo sapiens GN=NPTN PE=1 SV=2
881	SMAP_HUMAN	186	20378				0.9902	1.1187	2	Small acidic protein OS=Homo sapiens GN=SMAP PE=1 SV=1
882	BIEA_HUMAN	185	33693	**1.1570**	**1.0008**	2	1.1441	1.1321	3	Biliverdin reductase A OS=Homo sapiens GN=BLVRA PE=1 SV=2
883	LRP10_HUMAN	185	78027	0.6679	1.5087	2	1.2244	1.0847	2	Low-density lipoprotein receptor-related protein 10 OS=Homo sapiens GN=LRP1...
884	NRBP_HUMAN	185	60378	0.8153	1.1963	3	1.0048	1.0500	3	Nuclear receptor-binding protein OS=Homo sapiens GN=NRBP1 PE=1 SV=1
885	PSMD4_HUMAN	184	40940	**1.1684**	**1.0105**	4	**1.2834**	**1.1653**	4	26S proteasome non-ATPase regulatory subunit 4 OS=Homo sapiens GN=PSMD...
886	CNTP1_HUMAN	183	158220	0.9012	1.2492	2				Contactin-associated protein 1 OS=Homo sapiens GN=CNTNAP1 PE=1 SV=1
887	IDH3A_HUMAN	183	40023	1.1576	1.0703	2	1.0325	1.2972	4	Isocitrate

Matches (877 SSRA_HUMAN Mass: 32215 Score: 187 Matches: 6)

		z	Sequence	X	M/L	Std.Err	X	H/L	Std.Err	Fraction	Correlation	Intensity	Modification
1	+	3+	GTEDFIVESLDASFR	☑	1.2824	0.19464	☑	2.0323	0.15468	0.3200	0.9910	528181	
2	+	2+	GTEDFIVESLDASFR	☑	1.5968	0.10719	☑	2.0480	0.06616	0.1928	0.9696	7029959	

Mascot Distiller 2.6

MASCOT : *Mascot Distiller*

© 2016 Matrix Science

Here we have an example of this from a SILAC dataset – look at hit number 877. In Mascot Distiller 2.5 and earlier, this is reported as a protein ratio with two peptide matches, despite the matches being to the same peptide just at different precursor charge states. In Mascot Distiller 2.6 this is now being treated as a one hit wonder – and because of my settings, the protein ratio isn't being calculated or reported.