

Quantitation - MS1 based methods

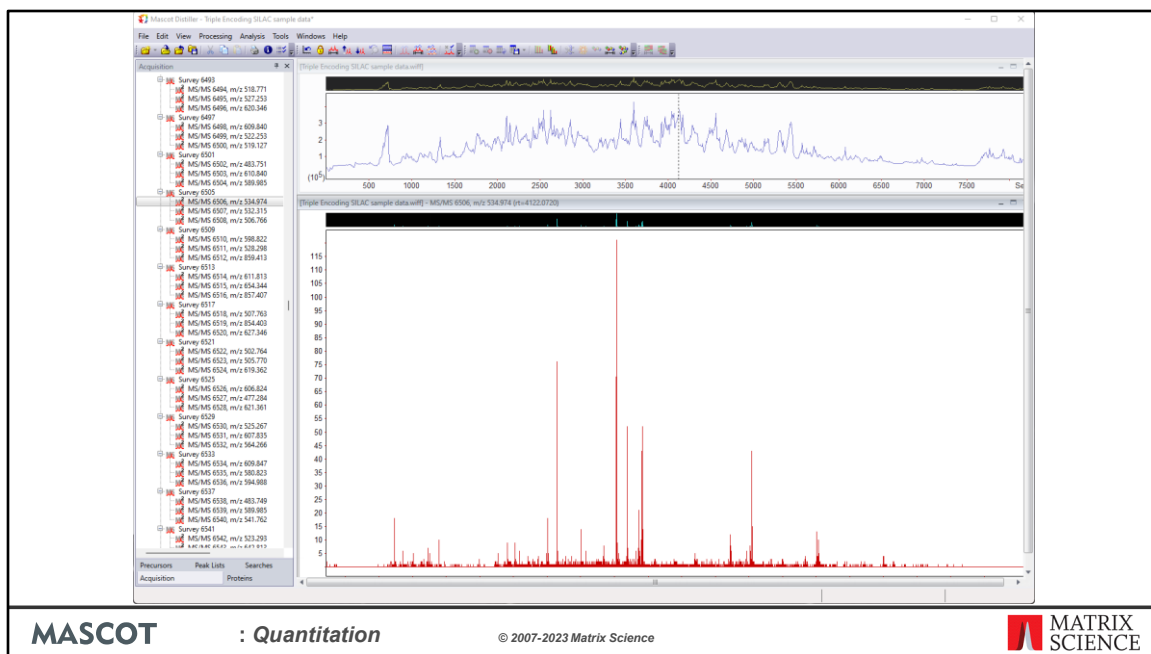
MASCOT

: *Quantitation*

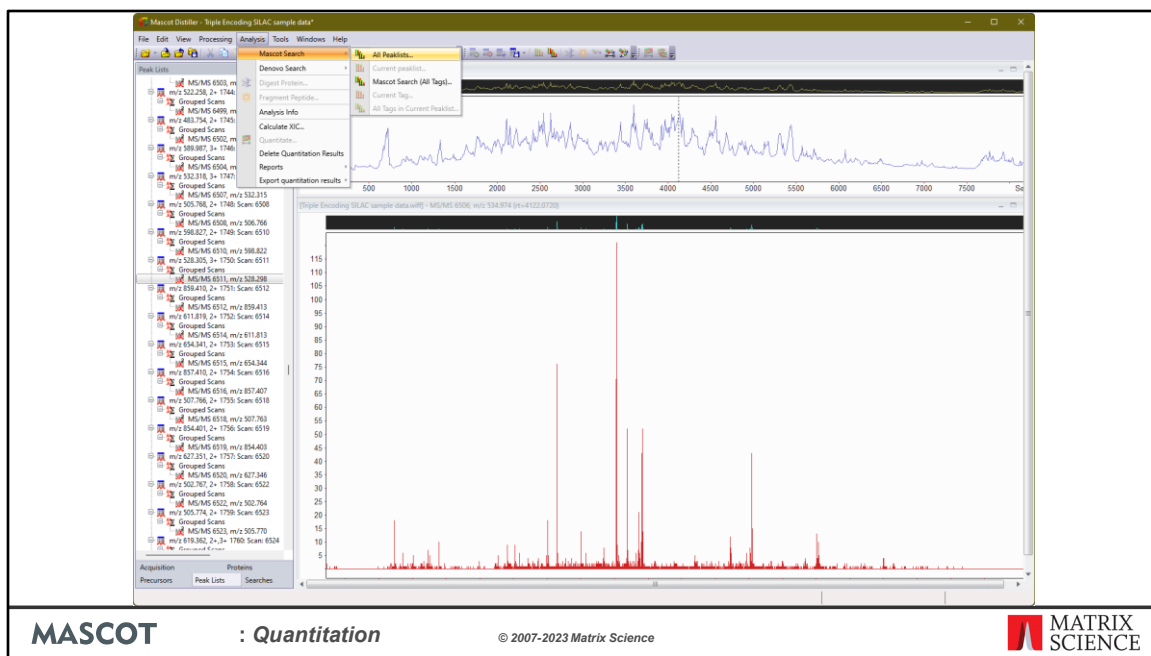
© 2007-2023 Matrix Science



In this presentation we will cover quantitation methods that require information from the raw file, such as peptide elution profile or peak intensities from the survey (MS1) scans. These MS1 based methods include the SILAC, N15 labelling and label free methods. You will need both Mascot Server and Mascot Distiller with the Search and Quantitation toolboxes.



I'm going to illustrate the workflow using a QStar sample data downloaded from the MSQuant website, Peter Mortensen et al. (2010). This is a SILAC sample with three components: unlabelled, Arg labelled with $^{13}\text{C}(6)$, and Arg labelled with $^{13}\text{C}(6)^{15}\text{N}(4)$. First, we open the Wiff file in Mascot Distiller, and process it into peak lists.



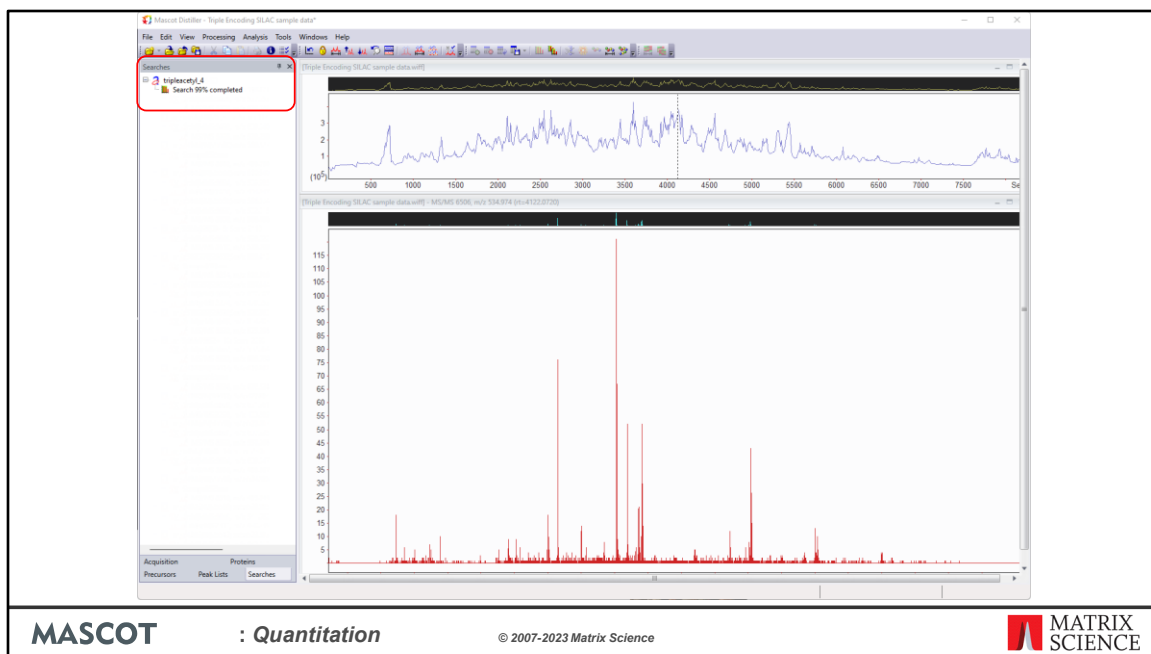
The peak lists are submitted to a Mascot Server to be searched.

MASCOT : Quantitation

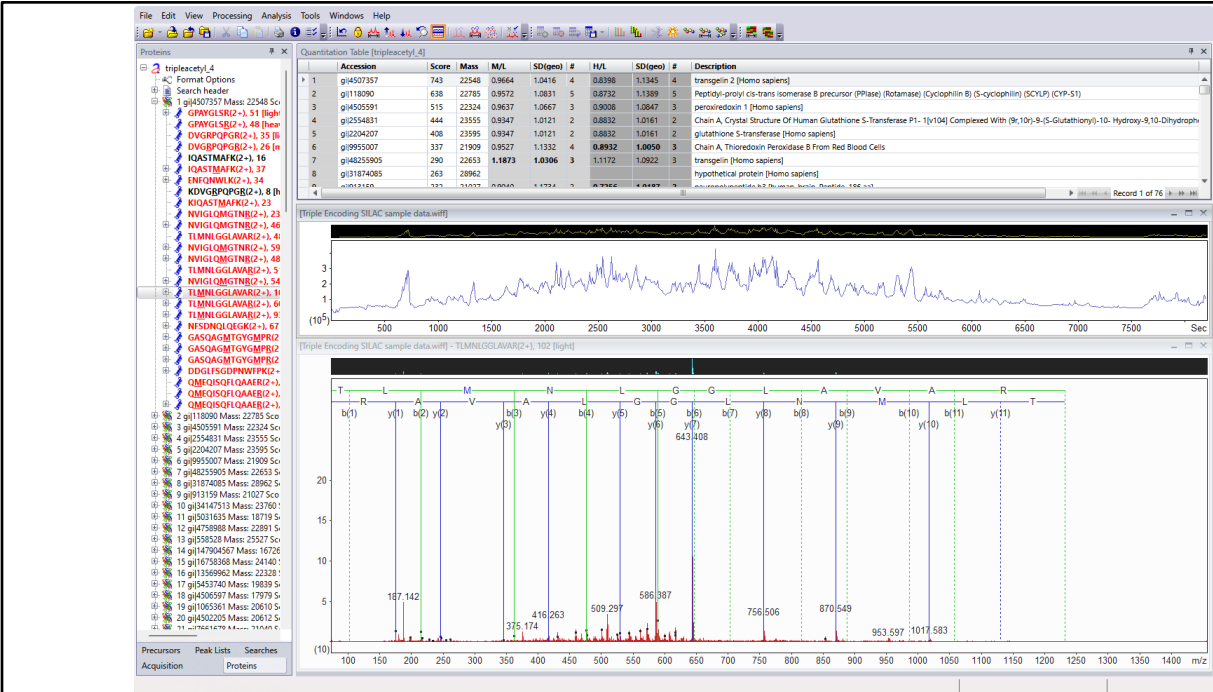
© 2007-2023 Matrix Science

MATRIX SCIENCE

There's an opportunity to tweak the search conditions, but most of the important settings, including the modifications, are embedded in the selected quantitation method.



We get progress reports while the search is running.

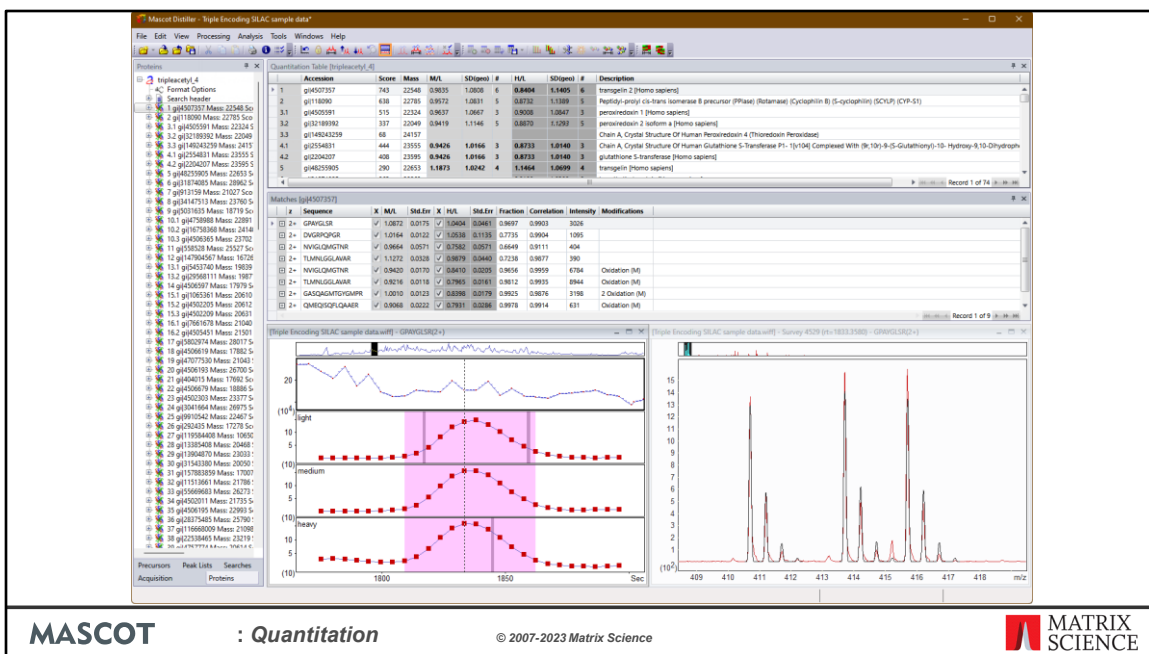


MASCOT : Quantitation

© 2007-2023 Matrix Science

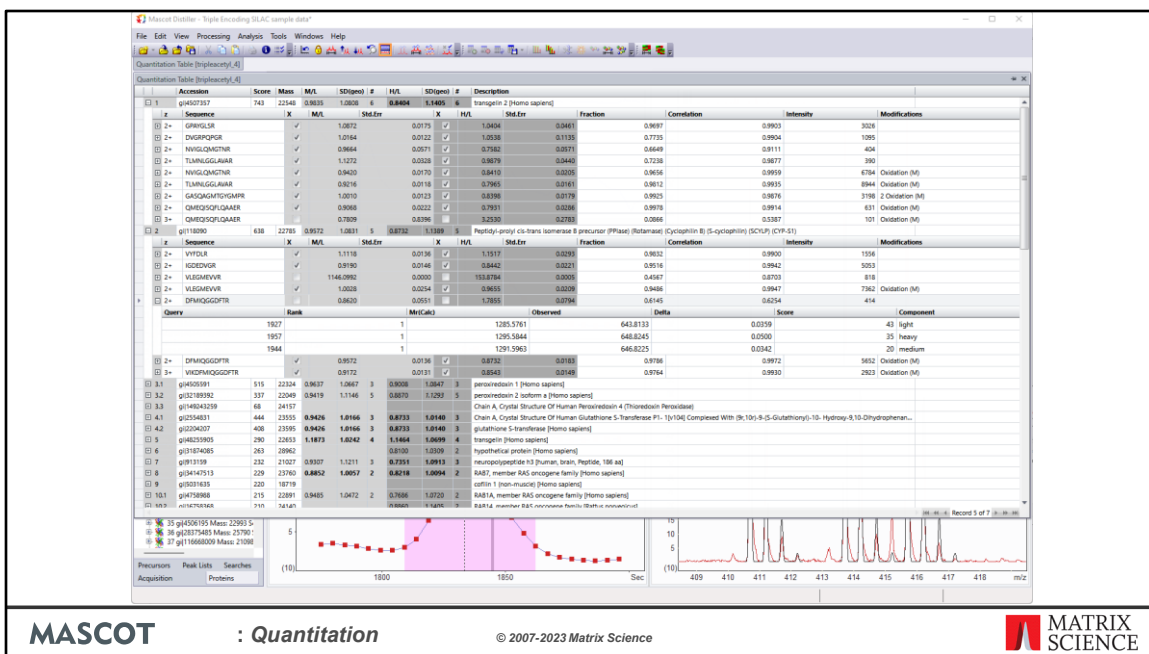


After a few minutes, the quantitation results are displayed as a table.



The quantitation results are displayed in a new grid control at the top right. Click on a protein to display a list of quantified peptides. Click on a peptide to display the TIC plus extracted ion chromatograms for the three components: light, medium, and heavy. The scan window at the bottom right shows the precursor region of the selected survey scan. The observed spectrum is in red. This is overlaid with black traces showing the isotope distributions calculated for the ratio being reported. You can make a visual judgement about the quality of the fit. The overlay can be turned on and off using a context menu.

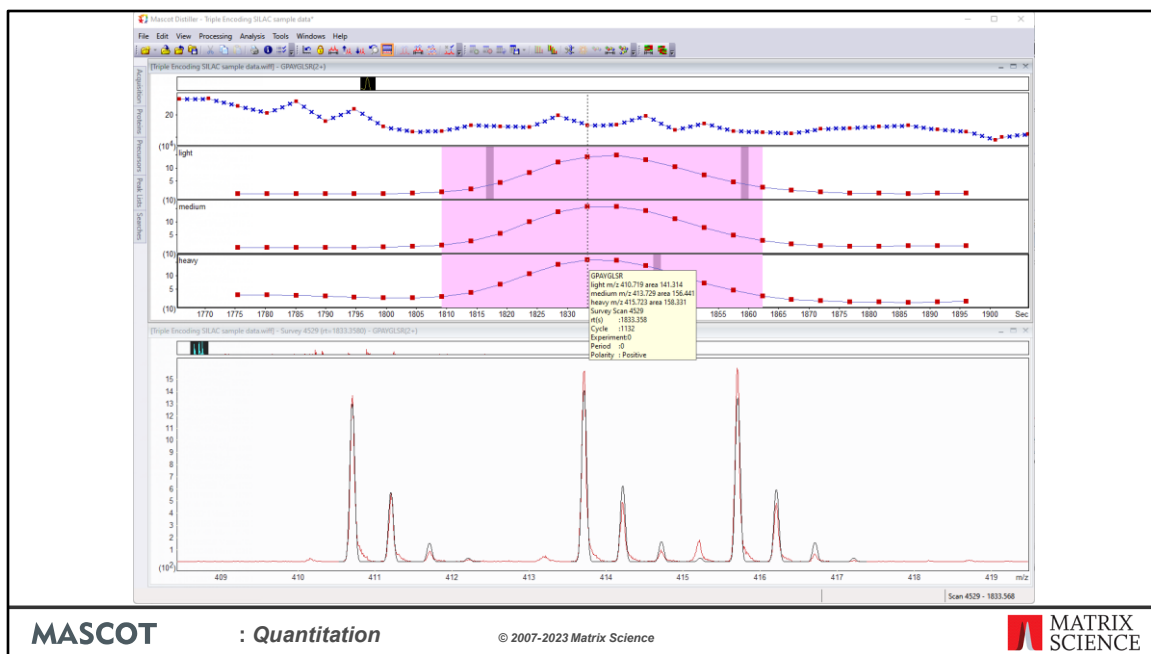
However big your screen, its always difficult to find enough room to display everything. To try and make best use of limited space, the grid and tree can be unpinned, so that they fly out when required and disappear when you move the mouse away.



Here, the grid has been unpinned and is being displayed over the top of the other windows. Two protein hits have been expanded to show individual peptides. One peptide has been expanded to show the corresponding Mascot matches. There's a lot going on here. If you want to study the numbers, this is the place to look. If you don't, then you can hide many of the columns to make the display simpler.

Because this is a three component experiment, two ratios are reported: M/L and H/L. These labels come from the method, and could equally well use sample IDs or whatever you prefer.

The checkboxes allow you to over-ride the decisions made by the software. If a checkbox is cleared, the peptide ratio is rejected and does not contribute to the protein ratio.

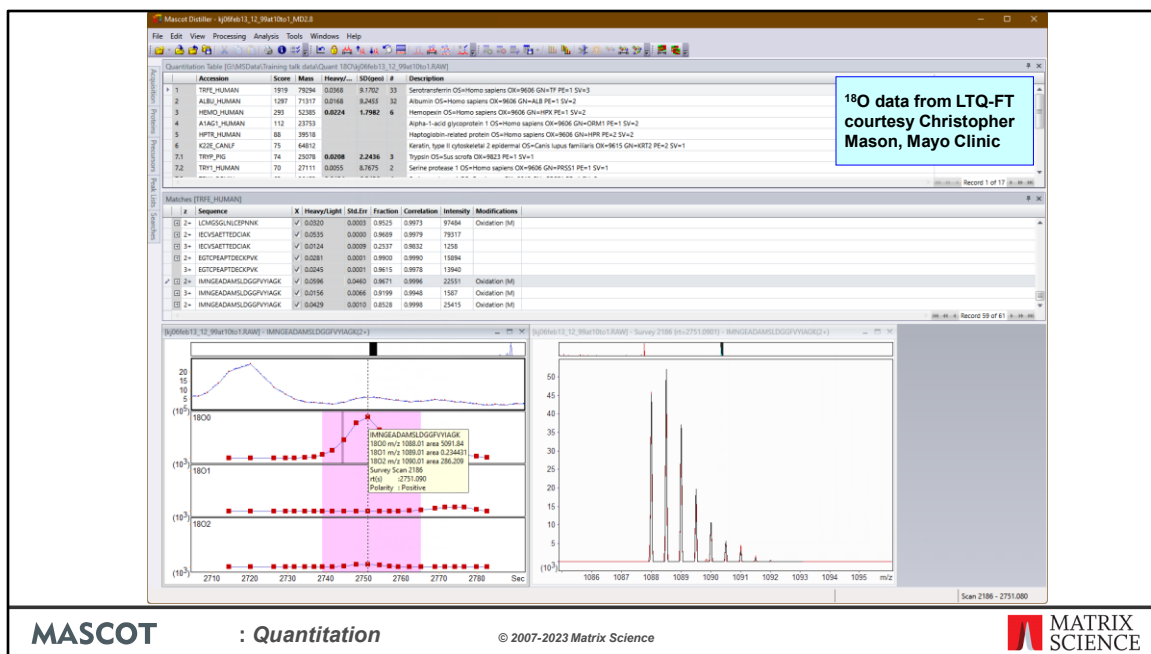


You can spend many happy hours devising different ways to arrange the windows.

With the tree and quant grid unpinned, we can see the chromatogram and scan windows more clearly. A tooltip for each scan provides mass and charge information.

The light purple region is the XIC peak. These are the scans that have been integrated to determine the ratios. The grey bands, which look dark purple when within the XIC peak, are the scans for which we have Mascot matches. In this particular case, there are matches for all three components. If there is no match for a component, the precursor is inferred from its mass and co-elution.

If isotope distributions overlap, this is accounted for in all calculations. In any individual scan, the fit between the experimental and calculated distributions is unlikely to be perfect. Even so, in this case, I hope you'll agree that we can immediately judge that the ratio isn't too bad and we can see there are no serious interferences or other problems in this particular scan.



**¹⁸O data from LTQ-FT
courtesy Christopher
Mason, Mayo Clinic**

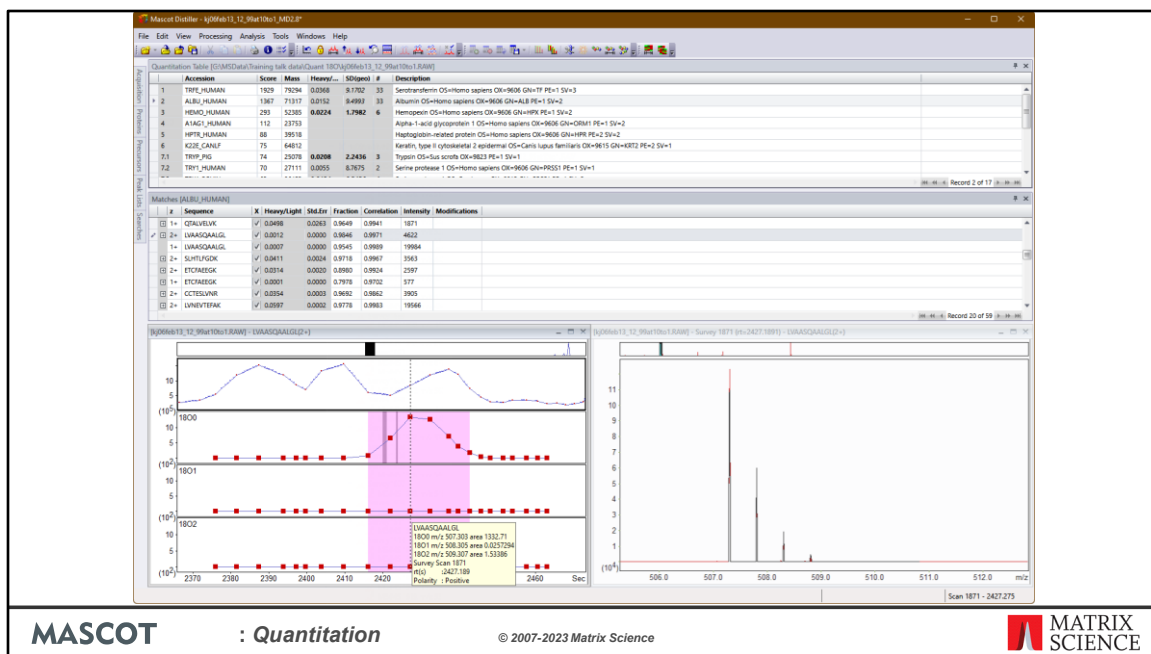
MASCOT

: Quantitation

© 2007-2023 Matrix Science

**MATRIX
SCIENCE**

This is O-18 data from an LTQ-FT, courtesy Christopher Mason, Mayo Clinic. The sample is part of a dilution series and this one is particularly tricky because the ratio is 10:1, light to heavy. Obviously, with only a 4 Da separation, the isotope distribution for the heavy is smoothly overlapped by the tail of the light distribution, particularly for larger peptides. At the protein level, the ratios come out around 0.08, which we think isn't too bad.



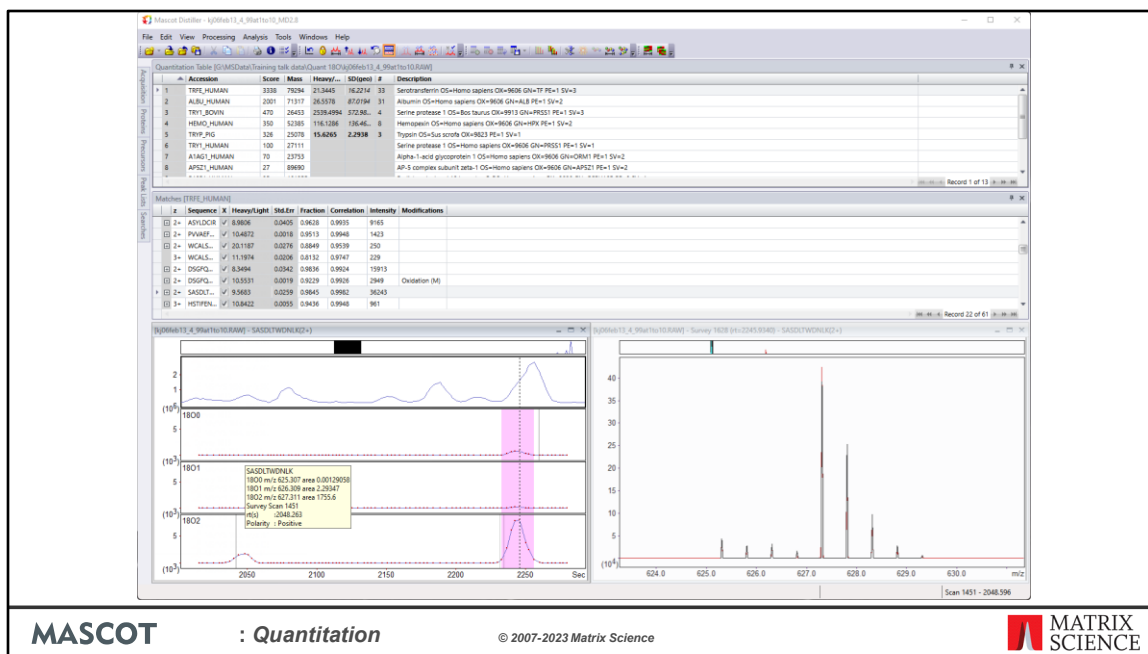
MASCOT

: Quantitation

© 2007-2023 Matrix Science

**MATRIX
SCIENCE**

This is the C-terminal peptide of BSA, which should be unlabeled, so reassuring to see the ratio is indeed close to zero. Note that to see the C-terminal peptide you need to edit the quantitation method as the default method uses the composition qualifier *[KR] to filter out all peptides that don't end with Lysine or Arginine and are hence not quantifiable.



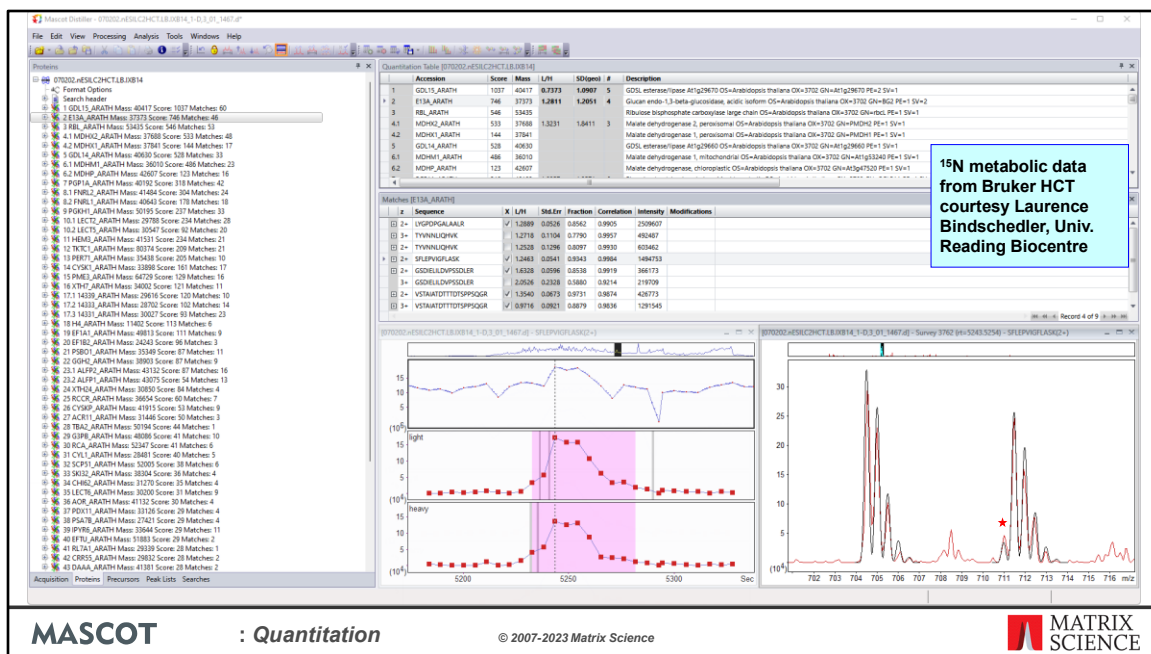
MASCOT

: Quantitation

© 2007-2023 Matrix Science

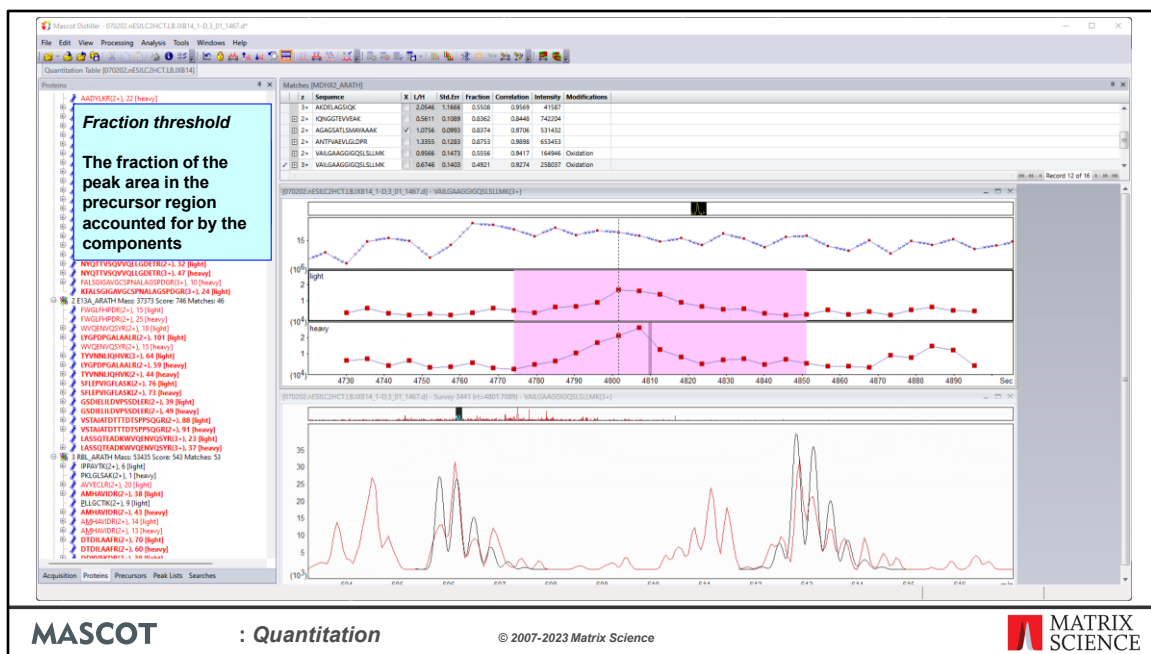
**MATRIX
SCIENCE**

Sometimes, XIC peak detection is a challenge. The starting point is set of scans for which we have Mascot matches. You can hardly see, but in this case, there is a grey band way out here for a match obtained some 3 minutes before the main peak came through. The XIC peak detection algorithm has decided that the bulk of the signal is within the much narrower region to the right, as shown by the purple highlight. This, by the way, is still O-18, but the sample is now 1:10 light to heavy.



Now some metabolic data, courtesy of Rainer Cramer's group at the Reading Biocentre. In the heavy component, all the nitrogens in the proteins have been replaced by N-15. The calculated distributions (black) fit reasonably well to the observed peaks (red). Notice the peak marked with a red star. This is due to the 1% isotope impurity.

For this particular ratio, there is evidence for some interference between the light and heavy precursors, and also something happening off to the right, but in general, this is a clean spectrum and can be expected to give a good ratio. We expect that most people will want the software to make this judgment, most of the time. In the quantitation grid, the figures in the columns headed standard error, fraction, and correlation are measures of the quality of the spectrum in the precursor region. The software compares these numbers with thresholds to decide whether to accept or reject a ratio. If you want, you can over-ride these decisions using the checkboxes. I'll describe briefly how each of these criteria work.

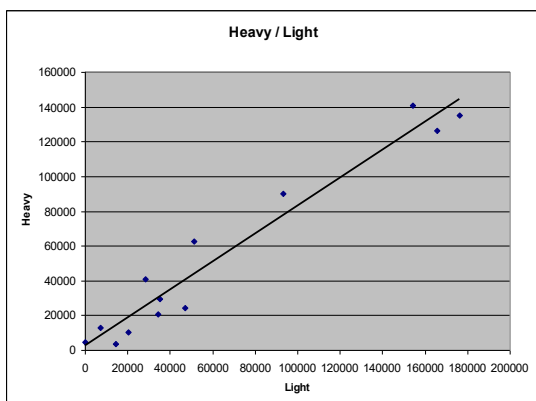


Fraction is the fraction of the peak area in the precursor region accounted for by the components. Here is a spectrum where there are a lot of interfering peaks. When we add up the areas, the expected precursors only account for 49% of the area, so the ratio is rejected. This threshold, like the others, is set as part of the quantitation method.



What about a case where the interfering peaks come right on top of the precursor peaks? A test on the matched fraction won't help. This is why we have a second test on the shape of the distribution. This uses the correlation coefficient between the predicted and observed precursor isotope distributions. Here is a case where a ratio fails the test, because the correlation coefficient is only 0.58.

Quantitation - Precursor



Each point represents the heavy and light intensities in a scan from the XIC peak. The straight line is a least squares fit

- The gradient is the best estimate of the ratio.
- Any background is corrected
- The standard error for the fit is a good measure of the reliability of the ratio

MASCOT

: Quantitation

© 2007-2023 Matrix Science

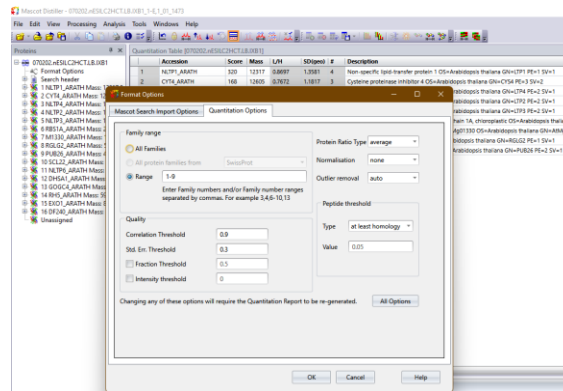
MATRIX
SCIENCE

Another column reports the estimated standard error for the calculated ratio. Each ratio comes from making a least squares fit to the component intensities from the scans in the XIC peak. Here, for example, each point represents the heavy and light intensities in one scan. The gradient of the fitted line is the best estimate of the ratio. The standard error for the fit is a good measure of the reliability of the ratio, and can simply be tested against another threshold.



Of course, as always, garbage in means garbage out. Here is a case where the raw data are centroids, not profiles. Although we have good strong MS/MS, when you look at the survey scans, there are no proper isotope distributions. The monoisotopic peaks are approximately in the right place, but the spacings to the isotope peaks are almost random. The charge state is 2+, and the calculated overlay shows what the distributions should look like. Presumably, there was something seriously wrong with the original peak picking. We simply cannot expect to get decent quantitative information out of data like this.

Quantitation options



MASCOT

: *Quantitation*

© 2007-2023 Matrix Science

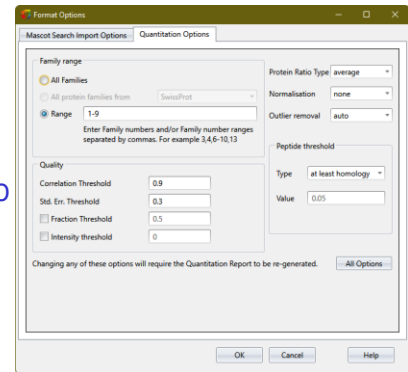
**MATRIX
SCIENCE**

The default settings for the Correlation, Standard Error, Fraction and Intensity thresholds are set in the quantitation method. They can be adjusted after quantitation has been completed through the Tools->Format options->Quantitation options dialog box. Making changes to the setting will cause Mascot Distiller to recalculate the results with the new filter settings applied.

Normalisation

When to use global normalisation

- Analysing equal total weights of protein from a complex mixture, e.g. cell lysate - **YES**
- Starting with equal numbers of cells - **NO**
- Isolating a sub-set of proteins by affinity methods - **NO**
- Looking at a synthetic dilution series - **NO**



MASCOT

: *Quantitation*

© 2007-2023 Matrix Science

**MATRIX
SCIENCE**

The Quantitation options also contain the normalisation setting.

Whether to calculate protein ratios from the average, median or weighted average of the set of peptide ratios is best decided by running some standards (e.g. a cell lysate spiked with varying amounts of a known protein) and seeing which gives the best accuracy and precision.

Normalisation is a way to reduce or eliminate systematic errors. In Mascot, you can normalise to one or more proteins or one or more peptide sequences. Normally, these will have been spiked into the sample for this purpose. You can also perform global normalisation by forcing the average or median ratio for all peptides to 1. If the average or median ratio is supposed to be 1, this is the smart thing to do. In other cases, it is the wrong thing to do. For example, if you are analysing a dilution series, where the ratio is supposed to be 3:1, you wouldn't want to force it to be 1:1.

Edit Quantitation Method: 15N Metabolic [MD]

Name: 15N Metabolic [MD] Description: 15N Metabolic, 99% labelling

Method Protocol **Component** Report Ratio Integration Quality Outliers Normalisation XML

Components: heavy New Copy Delete

Property	Value	Action
Component	heavy	

Modification groups Add Modification Group

Satellite group Add Satellite Group

Isotopes	Old	New	Delete
	N	15N	

Corrections	Type	Shift	Delete
	impurity	99.0	
	impurity	1.0	

Save changes Cancel

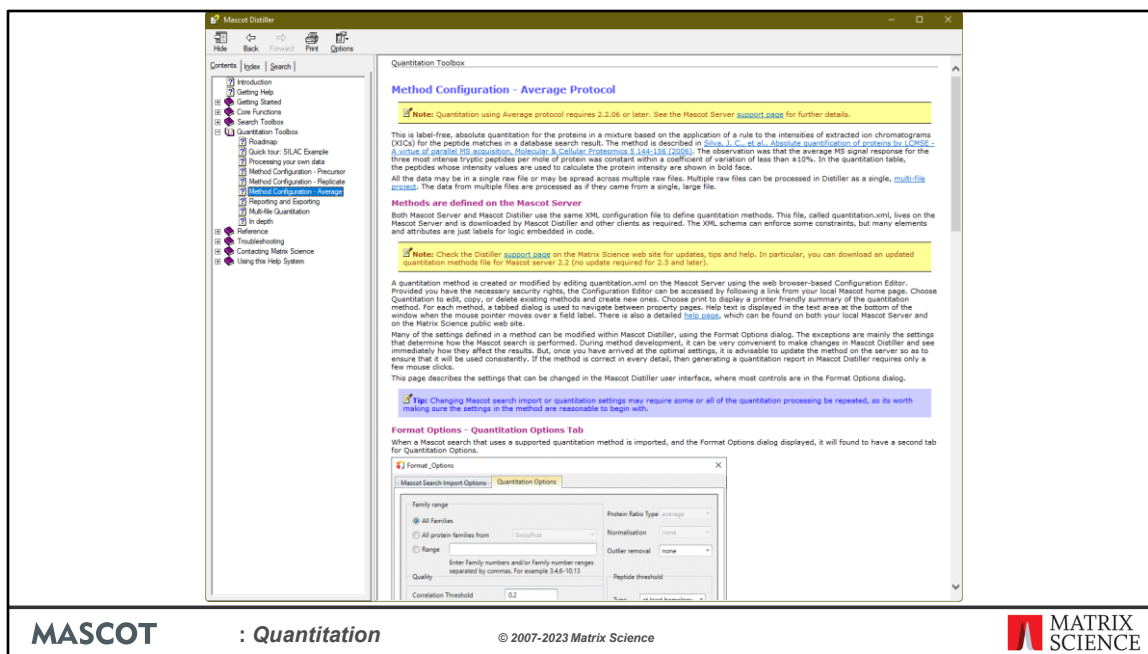
Isotope correction factor
 AB certificate=Percentage isotopic impurities for each reagent as tabulated on a Certificate of Analysis.
 averagine=Mass dependent isotope overlap. Isotope purity specified using element attribute of correction
 impurity=Isotope under-enrichment

Satellite peaks to higher m/z due to natural abundance of ^{13}C

Satellite peaks to lower m/z due to under-enrichment

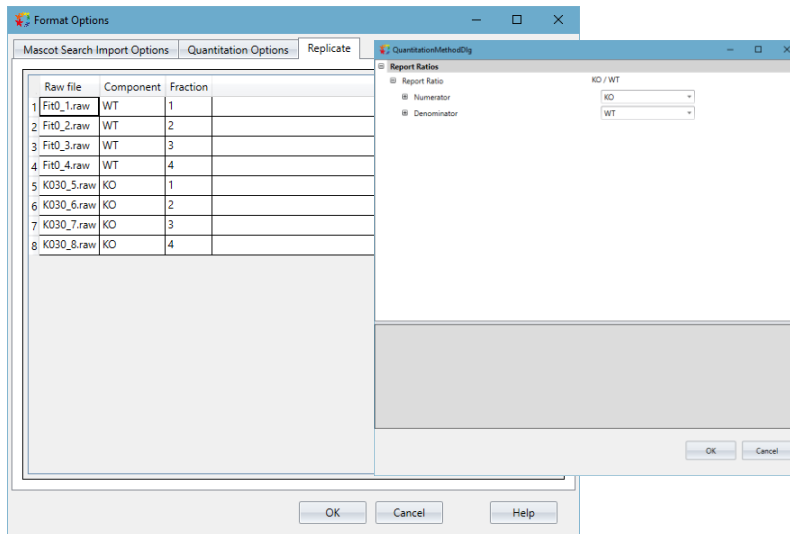
MASCOT : Quantitation © 2007-2023 Matrix Science **MATRIX SCIENCE**

One of the complications of any type of isotope labelling is isotope impurity. It is rarely possible to get 100% enrichment. In the Mascot quantitation schema, this is described by a correction element. An ‘impurity’ correction works “downwards”. That is, in this ^{18}O method, some of the intensity of peptides labelled with the ^{18}O label will appear at lower mass values because the heavy water is only 95% enriched. A second type of isotope correction, ‘averagine’, works “upwards”. This describes how some of the intensity will be found at higher mass values because of the natural abundances of heavy isotopes. An averagine correction only matters when the mass delta is small, as in the case of ^{18}O labelling.



In the interests of time, we'll skip over the Average protocol, but use this opportunity to point out that Distiller comes with comprehensive help. This page gives an overview of how Average works and details of all the configuration settings. There are similar pages for precursor and replicate.

Replicate



MASCOT

: Quantitation

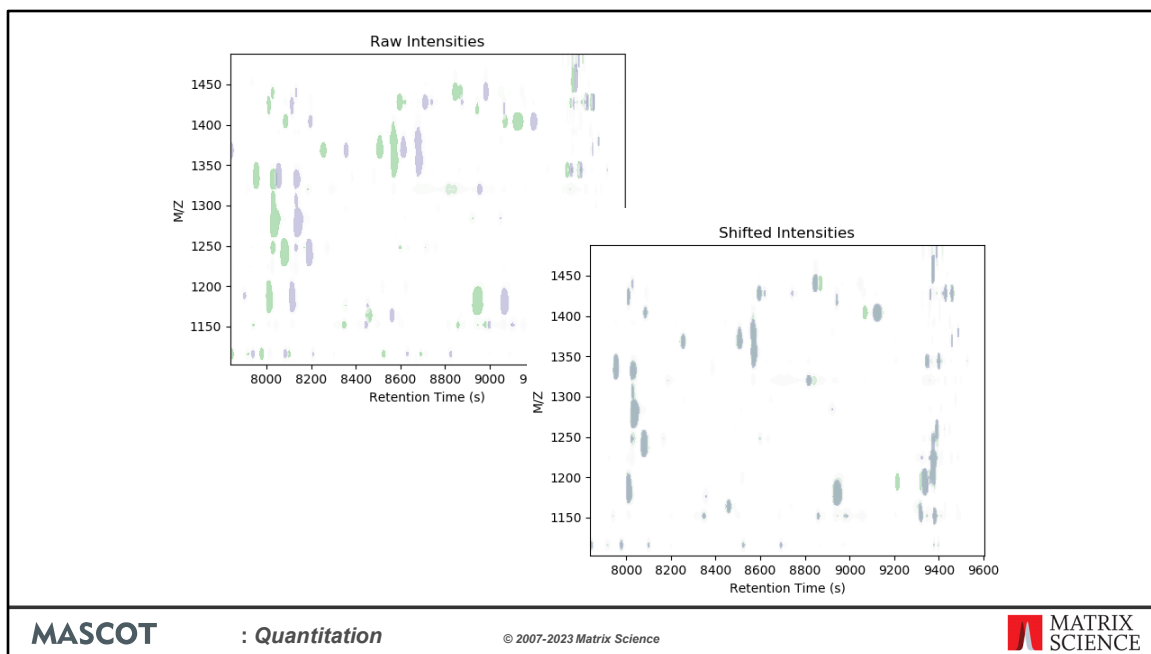
© 2007-2023 Matrix Science

**MATRIX
SCIENCE**

Replicate, you may remember, is 'label free'. Our implementation is identification driven, not feature driven. Distiller starts from the MS/MS data, imports the peptide matches from a Mascot search, then aligns them against a consensus generated by roughly aligning and combining the total ion chromatograms (TICs) of each raw file. A replicate project is always a multi-file project. It supports simple projects with one file per a sample or more complex data sets with multiple fractions per sample. As with precursor, you don't need to identify a peptide in every file.

Files are then aligned to the consensus chromatograms and XICs are calculated for peptides identified in one file but the other(s). For fractionated samples, multiple consensus are generated – one for each fraction. Files are then aligned to the consensus for their assigned fraction. This makes quantitation faster for these types of experiment as the system doesn't waste time looking for a peptide in all the files – it only looks for it in the fraction (or fractions) it was identified in.

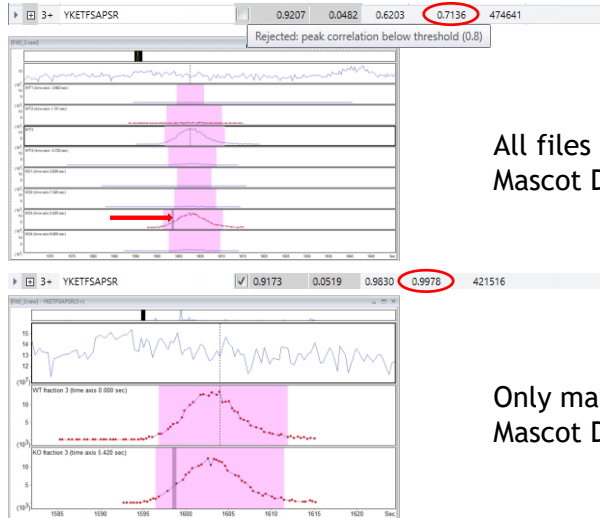
You can define custom ratios in the quantitation method or in Mascot Distiller directly.



Here we show the effect of the alignment, here we have MS1 features from two files in a label-free project plotted, with retention time on the X-axis and m/z on the Y-axis. Before alignment, we can see a clear shift in feature retention times between the two files. After time alignment, the common features between the two files are now clearly overlapping.

Note that this plot was created outside of Distiller for illustration, and is not currently available as a report.

GelC-MS Example: PXD029062



All files aligned together
Mascot Distiller 2.8.1 and earlier.

Only matching fractions aligned
Mascot Distiller 2.8.2 and later

MASCOT

: Quantitation

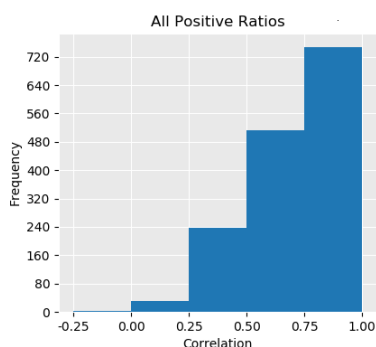
© 2007-2023 Matrix Science

**MATRIX
SCIENCE**

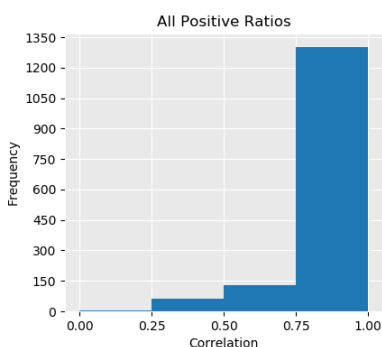
Here we have an example of how only aligning matching fractions is improving our results. This is a quantified peptides from the dataset where we've aligned all the files together. It's only been identified by a single MS/MS match in the KO sample, fraction 3. But here, Distiller has looked for peptide signal in all the files in the project – we have strong XICs from the WT and KO fraction 3 files, and essentially nothing from the other files. Searching in those files is wasted time.

The overall correlation for the observed and expected precursor isotope distributions for this peptide is 0.7136, because the junk from fractions 1,2 and 4 has dragged the correlation down. This means that the peptide has been rejected because the correlation is below our default threshold of 0.8 – despite actually having good XICs from the correct fraction.

In the second case, we have the same match from the dataset where we've only aligned files within their matching fractions. You can immediately see that we don't have XICs for the other fractions where the peptide wasn't identified. Our precursor isotope distribution correlation between observed and expected signal is now 0.9978 because we don't have any junk matching from the other fractions, and so the peptide has passed our correlation threshold quality filter.



Distiller 2.8.1
(align everything)



Distiller 2.8.2
(align fractions)

We can see the effect of this for the dataset as a whole using the “Quality” report in Distiller, which includes frequency histograms of various metrics. Here we can see that the number of peptides with a correlation of between 0.75 and 1.0 is increased from ~720 to ~1300 if we align within matching fractions only.

Quantitation: Ion Mobility Filtering

Bruker timsTOF data

Any precursor-based quantitation method

- E.g. LFQ, SILAC

Improves speed, accuracy and quality

Requires Mascot server 2.7 or later

- Must output precursor ion mobility in the peak lists

MASCOT

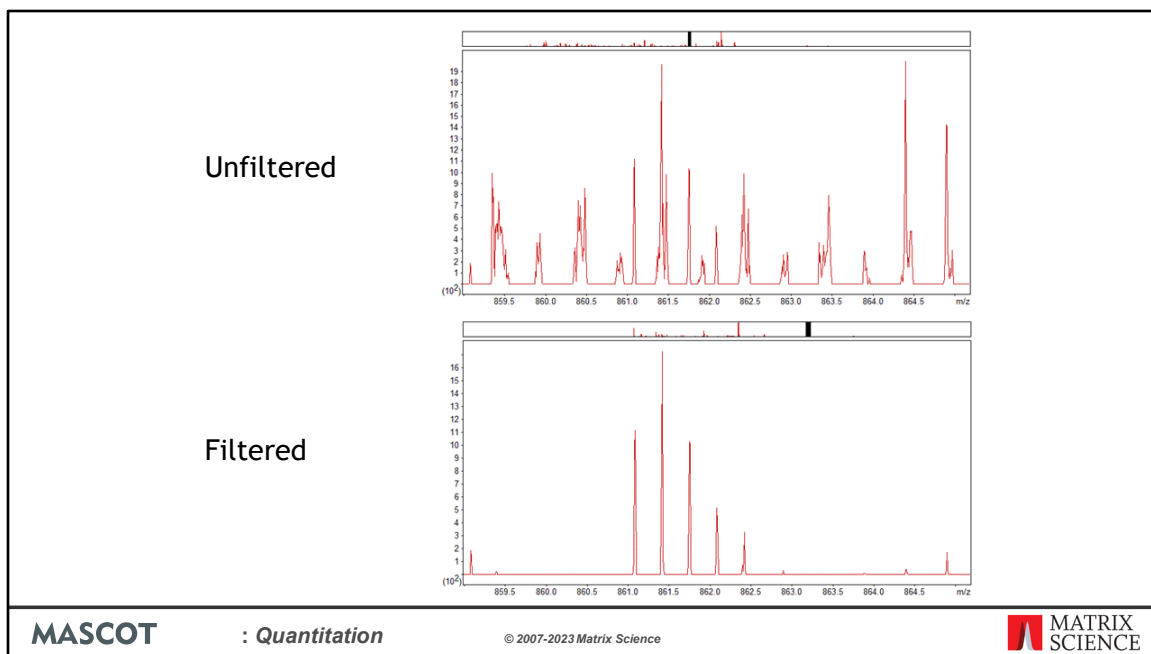
: *Quantitation*

© 2007-2023 Matrix Science

 **MATRIX
SCIENCE**

Mascot Distiller 2.8 can use the precursor ion mobility values for identified peptides as a filter on the precursor scans during the XIC peak detection steps of any precursor-based quantitation method, such as SILAC or Label Free. The aim of this is to clean up and improve the quality of the data in the target precursor regions, removing interfering signal, improving accuracy and giving more reliable quantitation results. For a given peptide match, an ion mobility range is calculated from the identified matches and the precursor ion mobility grouping tolerance defined in the processing options. Only MS1 signals from within the calculated ion mobility range are then used during XIC peak detection.

Ion mobility filtering requires that the data are searched using Mascot Server 2.7 or later, and that the precursor ion mobility values are included in the peak lists.



Here we can see an unfiltered scan from the precursor region of a peptide. This is a low intensity precursor and, as you can see, there is a lot of noise and interfering signal present.

With filtering enabled, the signal in the precursor area is much cleaner, with very little noise and interference, despite this being a low intensity precursor.

Quantitation: Ion Mobility Filtering

	Ion mobility filtering disabled	Ion mobility filtering enabled
Time taken (relative)	1	0.28
Median human protein ratio ^{a,b}	0.979	1.017
Number of human proteins quantified ^a	4458	4946
Median <i>E. coli</i> protein ratio ^{a,b}	4.21	4.3
Median <i>E. coli</i> proteins quantified ^a	718	821
Median peptide fraction value	0.487	0.686
Median peptide correlation value	0.912	0.959
Median protein ration sample size	5	6

MASCOT

: Quantitation

© 2007-2023 Matrix Science



**MATRIX
SCIENCE**

We took four files, representing two replicates, from the dataset used in the original PASEF paper for timsTOF. These samples consist of a background HeLa sample at a ratio of 1:1 and a spiked in *E. coli* protein digest with a reported median protein ratio of ~4.3. We then carried out quantitation on these samples with and without precursor ion mobility filtering enabled.

As you can see, enabling filtering significantly speeds up quantitation, because less data has to be processed. Median peptide fraction and correlation values also improved, showing the quality and signal to noise of the data are better with filtering enabled, allowing more peptide ratios to pass the quantitation quality thresholds.

The median number of quantitated peptide sequences used to calculate the protein ratios increased from 5 to 6, and more proteins with at least 2 distinct peptide sequences were quantified.

The median protein ratios for both human and *E. coli* proteins are also slightly closer to the expected ratios.

Quantitation talks

- K1. Quantitation - Introduction
- K2. Quantitation - MS2 based methods
- K3. Quantitation - MS1 based methods
- K4. Quantitation - Reporting

MASCOT

: *Quantitation*

© 2007-2023 Matrix Science



Please see the other quantitation presentations to learn about reporter ions, SILAC and label free quantitation as well as reporting formats for the results.

- K1. Quantitation – Introduction.
- K2. Quantitation - MS2 based methods. Quantitation methods that only require information available in the MS/MS peak list are supported in Mascot Server.
- K3. Quantitation - MS1 based methods. Methods that require additional information from the raw data file require Mascot Distiller + Quantitation Toolbox.
- K4. Quantitation – Reporting.