Quantitation

Quantitation was first introduced in Mascot 2.2. Our goal is to support all of the popular methodologies.

Quantitation - Overview

Protocol	Basis	Ratios	Examples
reporter	Specific reporter ion peaks within a single MS/MS spectrum	Inter-sample	iTRAQ, ExacTag, TMT, TMTpro
precursor	Extracted ion chromatograms for related precursors within a single dataset	Inter-sample	ICAT, SILAC, ^{18}O, ICPL, AQUA, Metabolic
multiplex (Neubert et. al.)	Pairs of sequence ion fragment peaks within a single MS/MS spectrum	Inter-sample	SILAC, ^{18}O
replicate	Extracted ion chromatograms for identical precursors across two or more datasets	Inter-sample	Label-free
empai (Ishihama et. al.)	Protein coverage from a database search result	Intra-sample	N/A
average (Silva et. al.)	Extracted ion chromatograms for selected peptides per protein within a single dataset	Intra-sample	N/A

To make this task manageable, we have classified the various approaches into a limited number of protocols. So far, we have identified 6 distinct protocols.

Reporter is quantitation based on the relative intensities of fragment peaks at fixed m/z values within an MS/MS spectrum. For example, iTRAQ or Tandem Mass Tags

Precursor is quantitation based on the relative intensities of extracted ion chromatograms (XICs) for precursors within a single data set. This is by far the most widely used approach, which can be used with any chemistry that creates a precursor mass shift. For example, 18O, AQUA, ICAT, ICPL, Metabolic, SILAC, etc., etc.

Multiplex is quantitation based on the relative intensities of sequence ion fragment peaks within an MS/MS spectrum. This is a novel approach, which can be used with any chemistry that labels one peptide terminus, creating a small mass shift, such as 18O or SILAC under certain conditions.

Replicate is label free quantitation based on the relative intensities of extracted ion chromatograms (XICs) for precursors in multiple data sets aligned using mass and elution time.

All these four methods are used to measure the relative abundance of a protein from sample to sample. For example, whether a particular protein is up or down regulated when an organism is stressed or diseased. The next two methods are used to estimate the relative abundances of different proteins within a single mixture.

emPAI is quantitation for the proteins in a mixture based on protein coverage by the peptide matches in a database search result.

Average is quantitation for the proteins in a mixture based on the application of a rule to the intensities of extracted ion chromatograms (XICs) for the peptide matches in a database search result. For example, the average intensity for the three strongest peptide matches per protein

The rows with a blue background are the protocols that implemented in the search engine, and don't require any additional software.

Quantitation - Overview

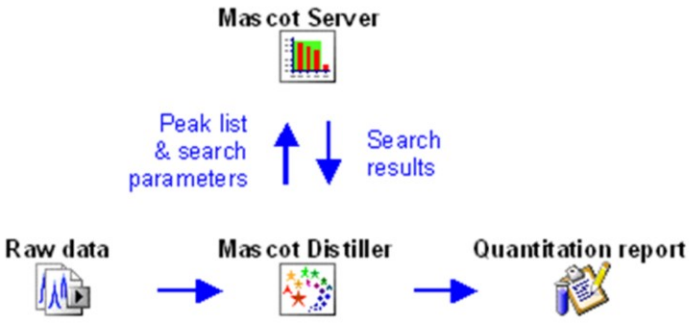
- Quantitation methods that only require information available in the MS/MS peak list are supported in Mascot Server
  - reporter
  - multiplex
  - emPAI
- Methods that require additional information from the raw data file require Mascot Distiller + Quantitation Toolbox
  - precursor
  - replicate
  - average

The common factor for these protocols is that all of the information required for quantitation is contained in the peak list.

The other three methods require additional information from the raw data file, either because it is necessary to integrate the elution profile of each peptide or because information is required for multiple peaks in the survey scan. These methods require that the raw data files are processed using Mascot Distiller.

Quantitation - Overview

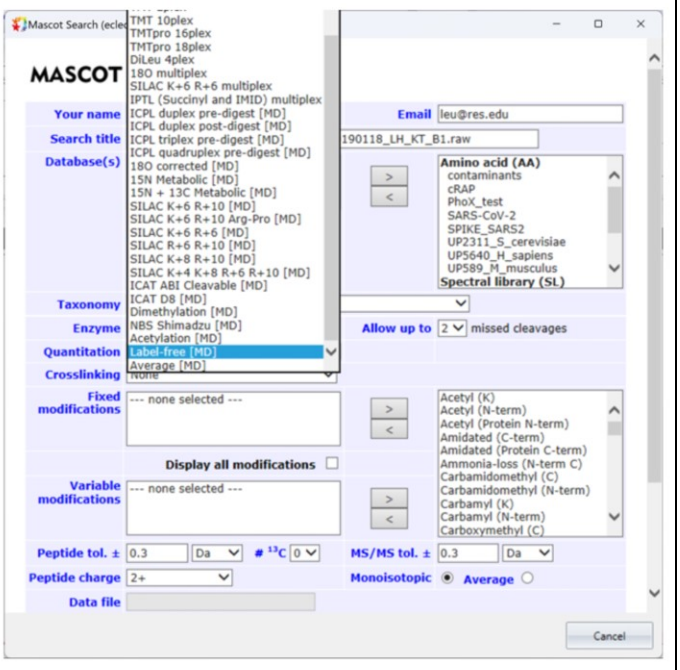
Workflow for methods that require additional information from the raw data file, (precursor, replicate, average)



For methods that require additional information from the raw data file, the workflow looks like this. The raw data file is processed in Distiller and the search submitted to Mascot. When the search is complete, the results are returned to Distiller. The quantitation report can then be generated in Mascot Distiller, which has access to both the Mascot search results and the raw data.

Quantitation

Named quantitation methods keep the search form uncluttered



We wanted to keep the user interface simple. Quantitation adds a huge number of choices and parameters, but there is no point in exposing all of these in the search form.

The approach we have chosen is encapsulate these choices and parameters into named quantitation methods. This means that the search form has just a single control.

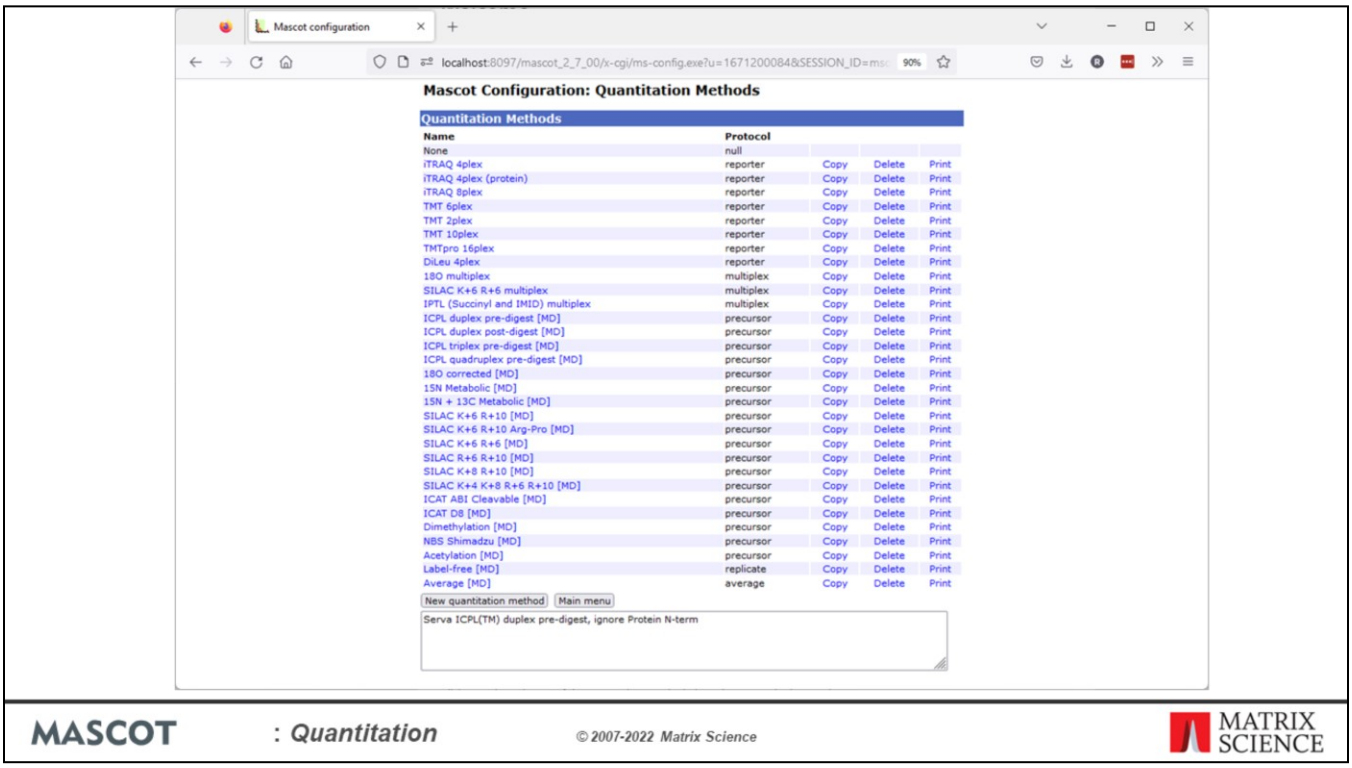
Methods that have [MD] at the end are the ones that require Mascot Distiller

Quantitation

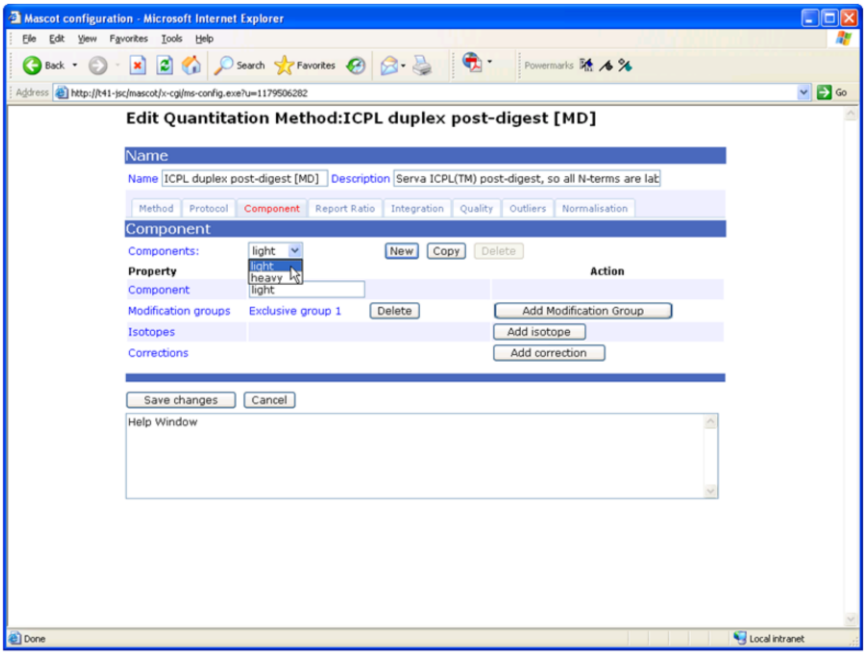
The quantitation methods are defined in a single XML configuration file

- quantitation.xml
- Browser based editor
- Add new methods as required
- Used by Mascot Server and Mascot Distiller

The configuration file that encapsulates the choices and parameters for each quantitation method is called quantitation.xml. This is an XML file, and there is a browser based editor for modifying methods and creating new ones. quantitation.xml lives on the Mascot server and is read by both the search engine and Mascot Distiller

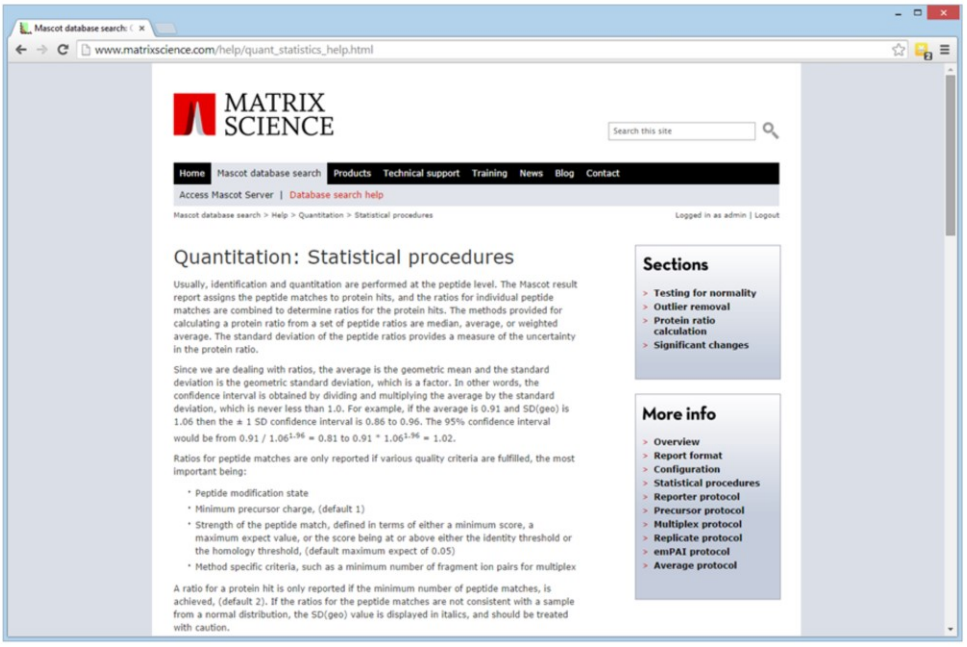


The browser-based Configuration Editor provides an interface to all the Mascot configuration files. In the case of quantitation, you can edit an existing method or make a copy of it as the basis for a new method



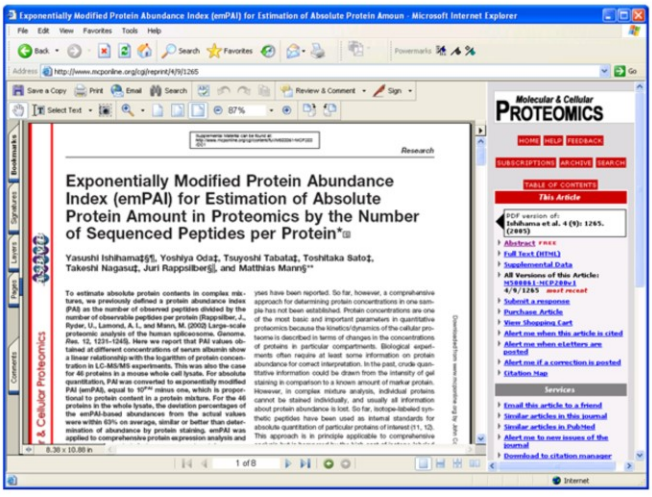
For each method, a tabbed dialog is used to navigate between property pages. In many cases, the property pages correspond to XML elements, but sometimes elements have been combined onto a single page or split across multiple pages so as to give a balanced layout.

Here, we can see a duplex ICPL method. The unlabelled and labelled components have been called heavy and light, but you are free to choose your own names so as to make the final report as clear as possible.



We have taken trouble to ensure that appropriate statistical procedures are correctly used. For example, we test that a set of peptide ratios is consistent with a normal distribution before rejecting outliers or reporting a standard deviation. Standard deviations are always geometric, because we are dealing with ratios that conform to a normal distribution in log space.

Quantitation - emPAI



emPAI quantitation offers approximate, label-free, relative quantitation of the proteins in a mixture based on protein coverage by the peptide matches in a database search result. This approach was developed by Ishihama and colleagues

Quantitation - emPAI

- Very simple

$$emPAI = 10^{\frac{N_{observed}}{N_{observable}}} - 1$$

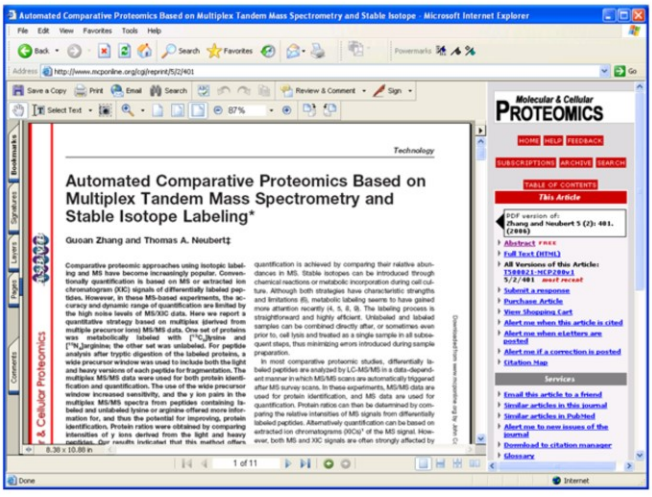
- Very approximate
  - Many assumptions in $N_{observed}$ and $N_{observable}$
- 'Always on'

1. PPB1_HUMAN **Mass:** 58259 **Score:** 452 **Queries matched:** 17 **emPAI:** 1.04
Alkaline phosphatase, placental type precursor (EC 3.1.3.1) (PLAP-1) (Regan isozyme) - Homo sapiens
Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
27	462.6807	923.3468	923.5116	-0.1649	0	33	0.25	1	R.FPYVALSK.T
41	517.1760	1032.3375	1032.5604	-0.2229	0	71	6.4e-05	1	R.GSSIFGLAPGK.A
62	564.6804	1127.3463	1127.5764	-0.2301	0	9	1.2e+02	1	R.GFFLFVEGGR.I

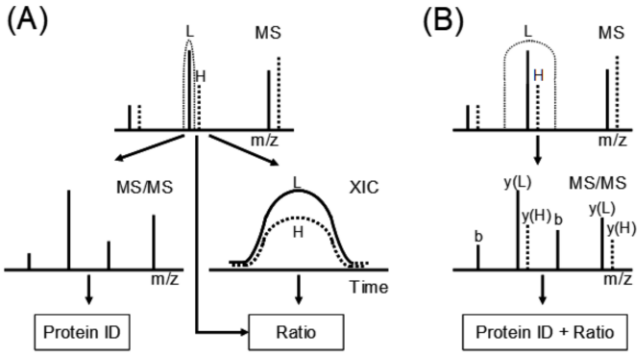
It is very simple. It is also very approximate, because there are many arbitrary assumptions in the way that the number of observed and observable peptides are calculated. Nevertheless, Ishihama's paper shows that it can be a useful guide to relative amounts. emPAI doesn't require a label or special data processing, so it is always reported in a standard Mascot results report, as long as the number of MS/MS spectra is at least 100

Quantitation - multiplex



Multiplex is quantitation based on the relative intensities of sequence ion fragment peaks within an MS/MS spectrum. This approach, developed Zhang and Neubert, can be used with any chemistry that labels one peptide terminus and has a reasonably small mass shift.

Quantitation - multiplex



This diagram, copied from the MCP paper, illustrates how it works. On the left, we have conventional quantitation; the 'precursor protocol' in Mascot terms. This requires the precursor intensity for each component to be integrated across its elution profile. In the case of the multiplex protocol, the MS1 transmission window is set wide enough to allow both components through simultaneously, giving a mixed MS/MS spectrum. The relative amounts can be measured from the sequence ions that include the labelled terminus. If the label is on the carboxy terminus, we see the ratios in the y ions.

Quantitation - multiplex

Requirements:

- Label confined to one peptide terminus
  e.g. ^{18}O, or SILAC at K or R with trypsin
- MS1 transmission window must be ~ flat over the label delta
- Heavy and light pair must be 'isolated' in survey scan
- Heavy and light must ~ co-elute
- Label must not affect fragmentation kinetics
- Tough to extend to more than 2 components.

The multiplex method has the potential to give excellent precision, because each ratio is represented by multiple sequence ion pairs. On the other hand, the ratio will only be accurate if several constraints are met.

Quantitation - multiplex

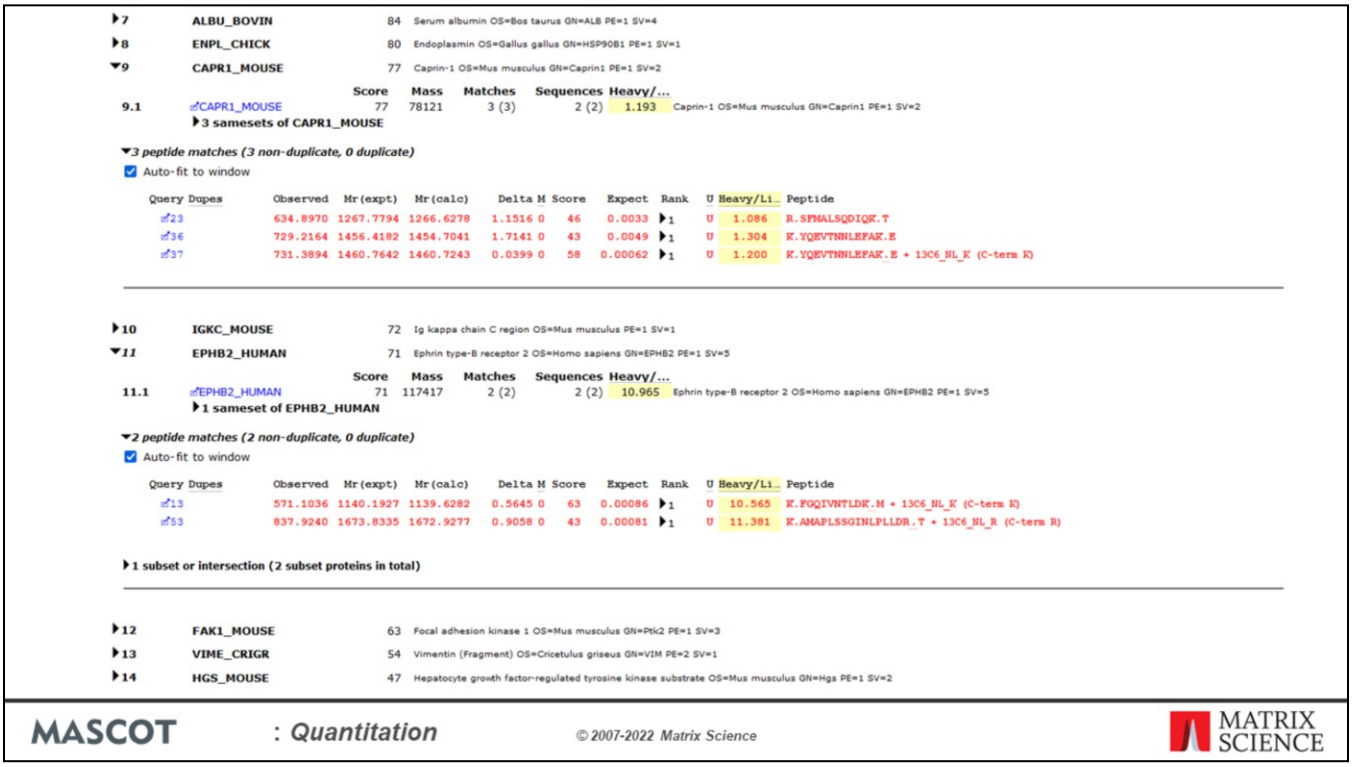
Isobaric Peptide Termini Labeling (IPTL):

- Koehler, C. J., et al., Isobaric Peptide Termini Labeling for MS/MS-Based Quantitative Proteomics, J. Proteome Research 8 4333-4341 (2009)
- Label both termini
- Heavy and light have equal and opposite shifts, e.g.

  Component 1:
  Succinyl d0 at the N-term and IMID d4 at C-term

  Component 2:
  Succinyl d4 at the N-term and IMID d0 at C-term

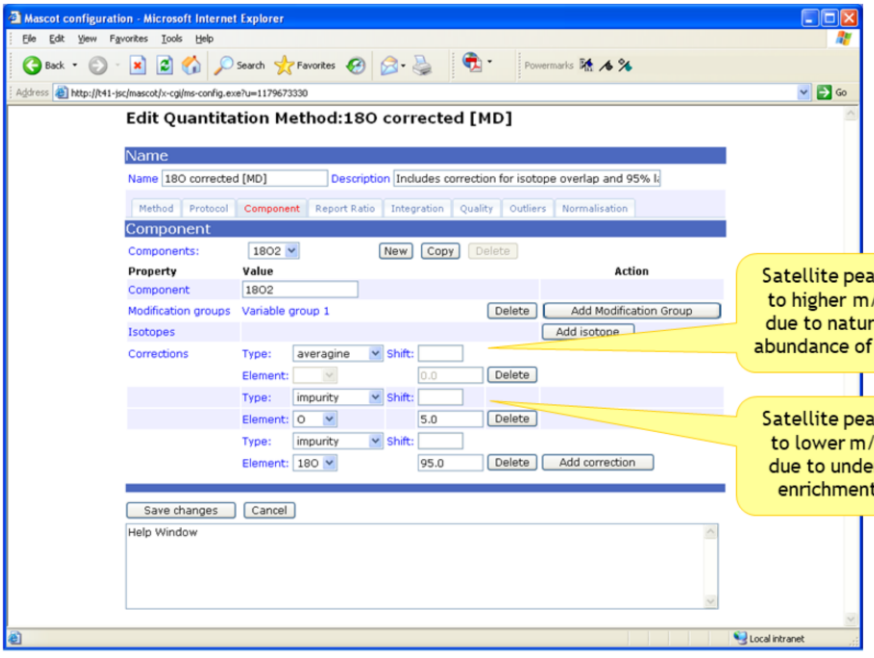
Isobaric Peptide Termini Labeling (IPTL) is a recent improvement to multiplex. This labels both termini and the difference between the two components is a mass increase at one terminus exactly balanced by a mass decrease at the other. Having isobaric precursors removes the requirement for the transmission window between MS1 and MS2 to be wide enough to accommodate the mass shift introduced by the label.



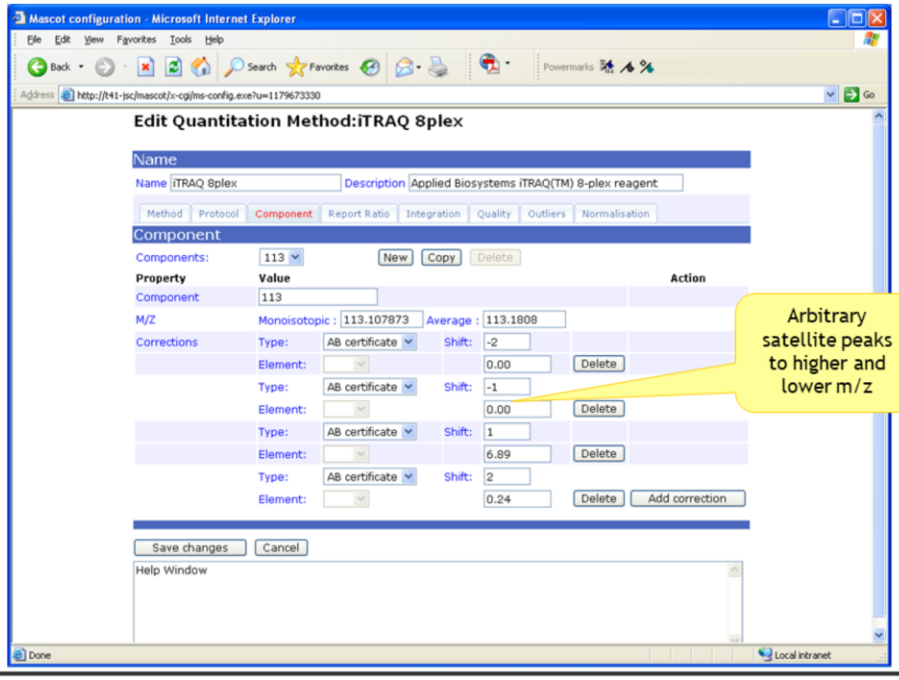
This is an example of Multiplex using a dataset courtesy of Zhang and Neubert. The instrument was an ion trap and the label is 13C(6) SILAC on K and R.

#	b	b^{++}	b^*	b^{*++}	b^0	b^{0++}	Seq.	y	y^{++}	y^*	y^{*++}	y^0	y^{0++}	#
1	148.0757	74.5415					F							10
2	205.0972	103.0522					G	993.5671	497.2872	976.5405	488.7739	975.5565	488.2819	9
3	333.1557	167.0815	316.1292	158.5682			Q	936.5456	468.7764	919.5191	460.2632	918.5350	459.7712	8

We can see that the heavy component has been strongly up-regulated in this peptide from human ephrin

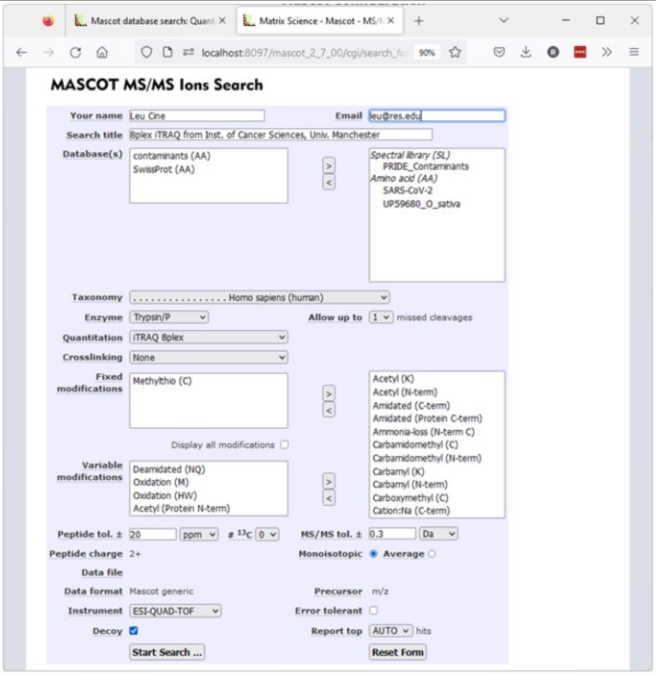


One of the complications of any type of isotope labelling is isotope impurity. It is rarely possible to get 100% enrichment. In the Mascot quantitation schema, this is described by a correction element. An ‘impurity’ correction works "downwards". That is, in this 18O method, some of the intensity of peptides labelled with the 18O 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 18O labelling.

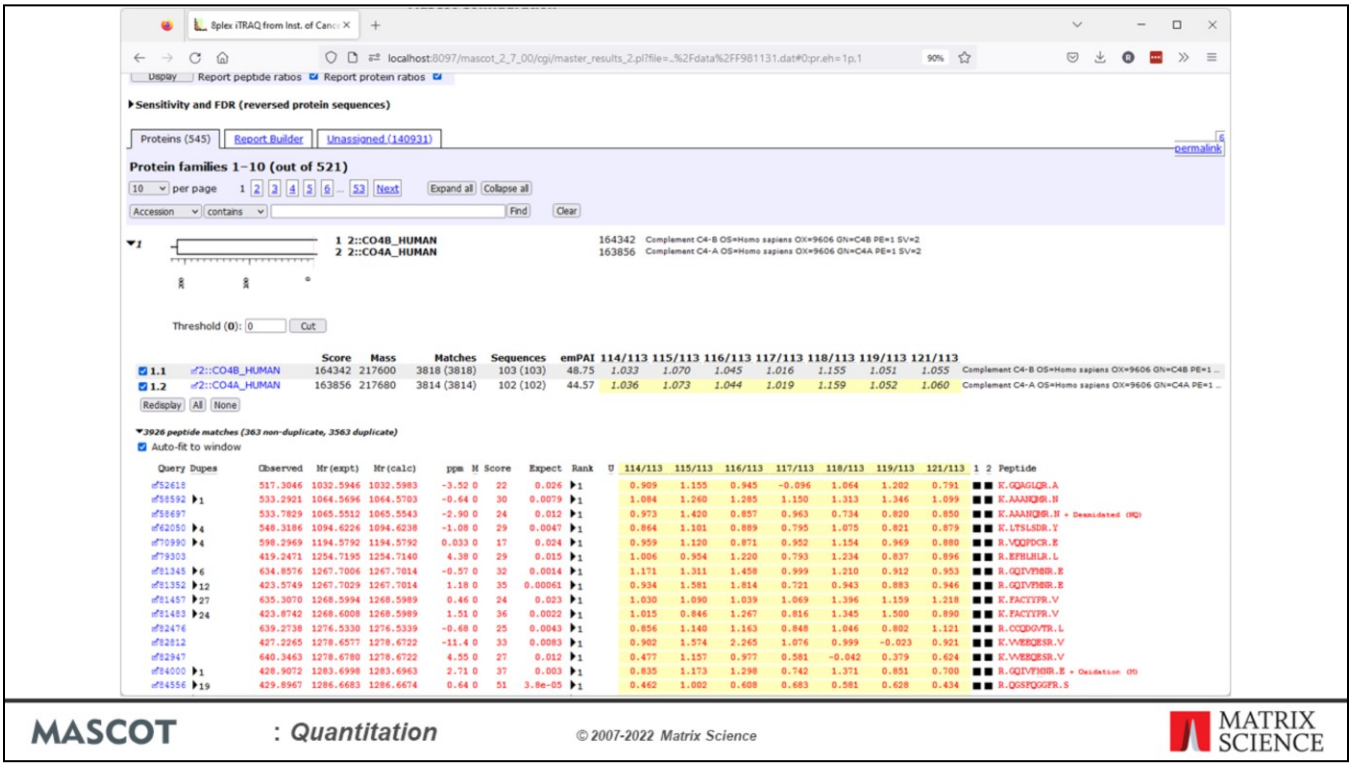


A third type of isotope correction is used in iTRAQ and TMT, where the correction factors are obtained experimentally, by analysing the isolated reagents. This combines both upward and downward corrections for labels which have complex, multi-isotope compositions

Quantitation - Reporter

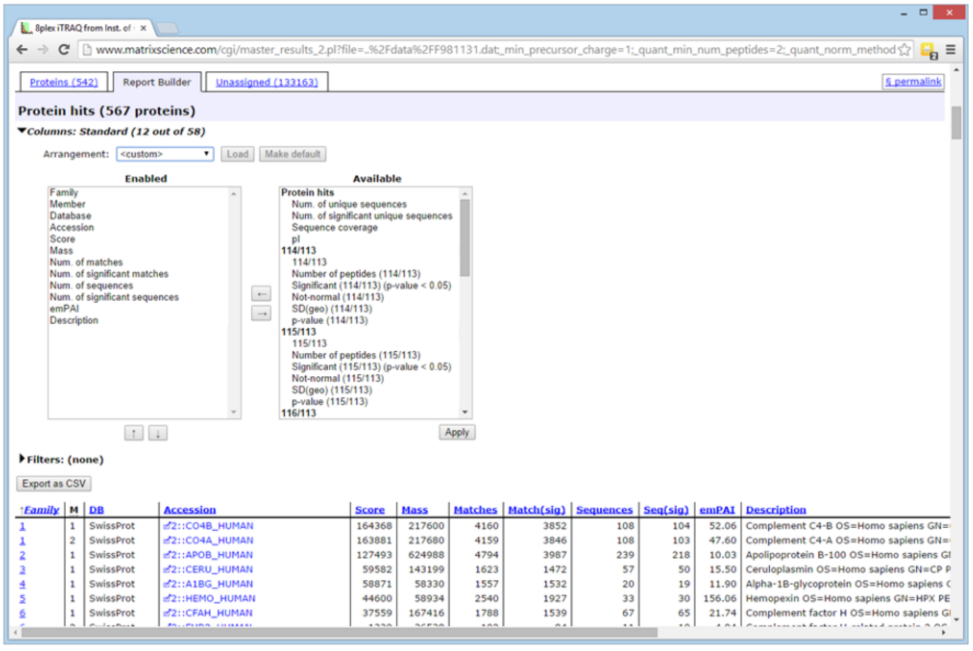


For reporter ion quantitation everything happens on the search engine as the peak list contains all the information required for the quantitation. Open the search form. We choose an appropriate quantitation method. We don't need to specify the iTRAQ modifications because these are pre-defined in the quantitation method. Submit the search…



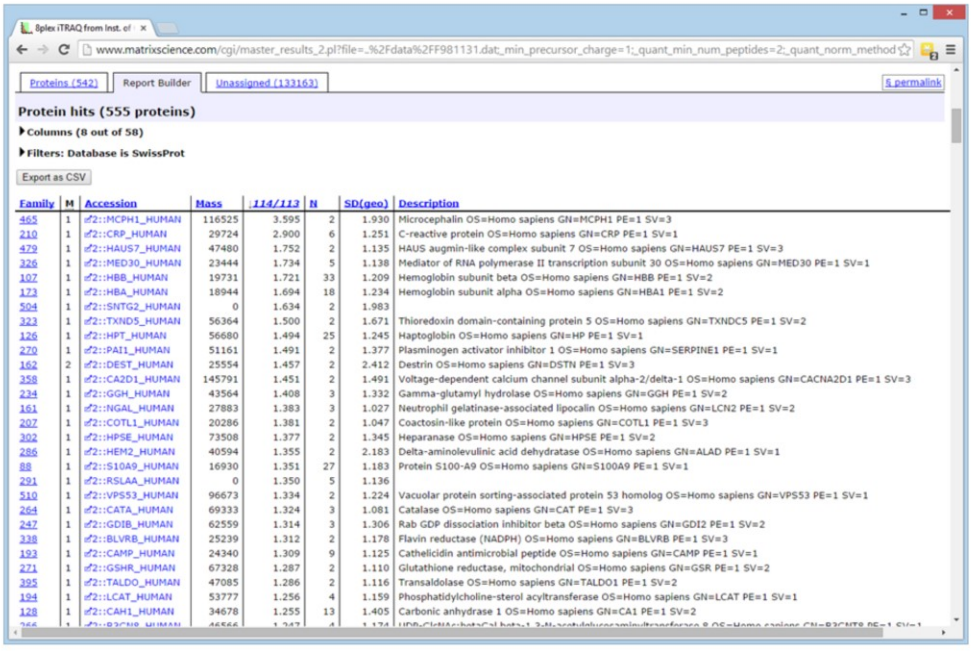
Unless it is a very small dataset, the result report will default to the Protein Family Summary. This example is 8plex iTRAQ. In the Proteins tab, you can toggle display of the ratios for proteins and for individual peptides. For the protein ratios, a tooltip shows the peptide ratio count, the geometric standard deviation, and the p-value for the ratio being different from 1

Which ratios are displayed and how they are named is specified in the quantitation method. You could edit the method to report different pairs, e.g. 115/114 and 117/116, or something more complex, like ratios to the sum of all four channels. Note that you can't do this if you are using our public web site, because this is a shared resource, so you don't have access to the configuration editor.



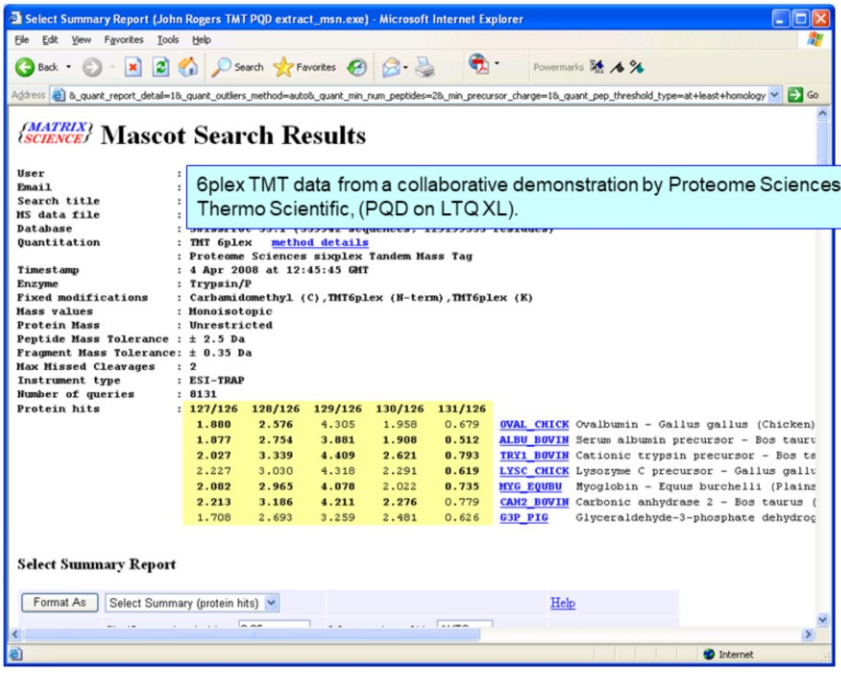
MASCOT : Quantitation © 2007-2022 Matrix Science MATRIX SCIENCE

We go to the report builder tab to configure a tabular report covering all the proteins of interest. You can select and re-order the columns, apply filters, and sort the rows.



Family	M	Accession	Mass	↓114/113	N	SD(geo)	Description
465	1	♂2::MCPH1_HUMAN	116525	3.595	2	1.930	Microcephalin OS=Homo sapiens GN=MCPH1 PE=1 SV=3
210	1	♂2::CRP_HUMAN	29724	2.900	6	1.251	C-reactive protein OS=Homo sapiens GN=CRP PE=1 SV=1
479	1	♂2::HAUS7_HUMAN	47480	1.752	2	1.135	HAUS augmin-like complex subunit 7 OS=Homo sapiens GN=HAUS7 PE=1 SV=3
326	1	♂2::MED30_HUMAN	23444	1.734	5	1.138	Mediator of RNA polymerase II transcription subunit 30 OS=Homo sapiens GN=MED30 PE=1 SV=1
107	1	♂2::HBB_HUMAN	19731	1.721	33	1.209	Hemoglobin subunit beta OS=Homo sapiens GN=HBB PE=1 SV=2
173	1	♂2::HBA_HUMAN	18944	1.694	18	1.234	Hemoglobin subunit alpha OS=Homo sapiens GN=HBA1 PE=1 SV=2
504	1	♂2::SNTG2_HUMAN	0	1.634	2	1.983	
323	1	♂2::TXND5_HUMAN	56364	1.500	2	1.671	Thioredoxin domain-containing protein 5 OS=Homo sapiens GN=TXNDC5 PE=1 SV=2
126	1	♂2::HPT_HUMAN	56680	1.494	25	1.245	Haptoglobin OS=Homo sapiens GN=HP PE=1 SV=1
270	1	♂2::PAI1_HUMAN	51161	1.491	2	1.377	Plasminogen activator inhibitor 1 OS=Homo sapiens GN=SERPINE1 PE=1 SV=1
162	2	♂2::DEST_HUMAN	25554	1.457	2	2.412	Destrin OS=Homo sapiens GN=DSTN PE=1 SV=3
358	1	♂2::CA2D1_HUMAN	145791	1.451	2	1.491	Voltage-dependent calcium channel subunit alpha-2/delta-1 OS=Homo sapiens GN=CACNA2D1 PE=1 SV=3
234	1	♂2::GGH_HUMAN	43564	1.408	3	1.332	Gamma-glutamyl hydrolase OS=Homo sapiens GN=GGH PE=1 SV=2
161	1	♂2::NGAL_HUMAN	27883	1.383	3	1.027	Neutrophil gelatinase-associated lipocalin OS=Homo sapiens GN=LCN2 PE=1 SV=2
207	1	♂2::COTL1_HUMAN	20286	1.381	2	1.047	Coactosin-like protein OS=Homo sapiens GN=COTL1 PE=1 SV=3
302	1	♂2::HPSE_HUMAN	73508	1.377	2	1.345	Heparanase OS=Homo sapiens GN=HPSE PE=1 SV=2
286	1	♂2::HEM2_HUMAN	40594	1.355	2	2.183	Delta-aminolevulinic acid dehydratase OS=Homo sapiens GN=ALAD PE=1 SV=1
88	1	♂2::S10A9_HUMAN	16930	1.351	27	1.183	Protein S100-A9 OS=Homo sapiens GN=S100A9 PE=1 SV=1
291	1	♂2::RSLAA_HUMAN	0	1.350	5	1.136	
510	1	♂2::VPS53_HUMAN	96673	1.334	2	1.224	Vacuolar protein sorting-associated protein 53 homolog OS=Homo sapiens GN=VPS53 PE=1 SV=1
264	1	♂2::CATA_HUMAN	69333	1.324	3	1.081	Catalase OS=Homo sapiens GN=CAT PE=1 SV=3
247	1	♂2::GDIB_HUMAN	62559	1.314	3	1.306	Rab GDP dissociation inhibitor beta OS=Homo sapiens GN=GDI2 PE=1 SV=2
338	1	♂2::BLVRB_HUMAN	25239	1.312	2	1.178	Flavin reductase (NADPH) OS=Homo sapiens GN=BLVRB PE=1 SV=3
193	1	♂2::CAMP_HUMAN	24340	1.309	9	1.125	Cathelicidin antimicrobial peptide OS=Homo sapiens GN=CAMP PE=1 SV=1
271	1	♂2::GSHR_HUMAN	67328	1.287	2	1.110	Glutathione reductase, mitochondrial OS=Homo sapiens GN=GSR PE=1 SV=2
395	1	♂2::TALDO_HUMAN	47085	1.286	2	1.116	Transaldolase OS=Homo sapiens GN=TALDO1 PE=1 SV=2
194	1	♂2::LCAT_HUMAN	53777	1.256	4	1.159	Phosphatidylcholine-sterol acyltransferase OS=Homo sapiens GN=LCAT PE=1 SV=1
128	1	♂2::CAH1_HUMAN	34678	1.255	13	1.405	Carbonic anhydrase 1 OS=Homo sapiens GN=CA1 PE=1 SV=2

As an example, maybe we want to list proteins with the largest fold change for 114/113 after excluding contaminants. The table has been sorted on descending 114/113.



The other widely used reporter chemistry is TMT from Thermo. This slide illustrates the appearance of the Select Summary, which is the default for small datasets

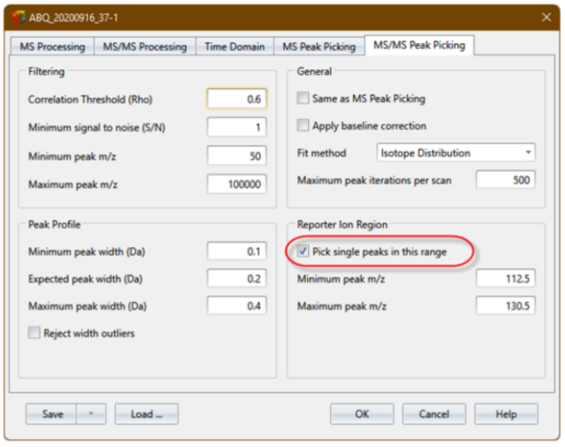
Peak picking

- Reporter ions are not peptides!
- Cannot use conventional de-isotoping
- Unless your peak picking software has support for a “reporter ion window”, turn off de-isotoping

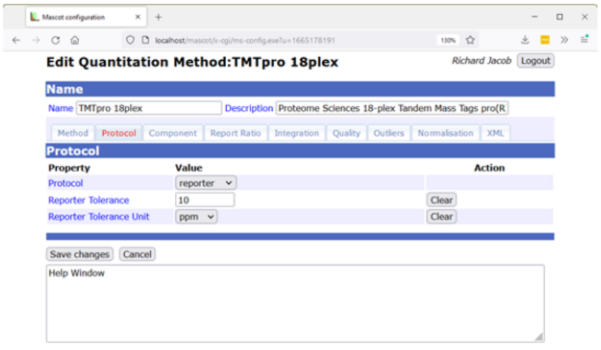
For the reporter protocol, i.e. iTRAQ or TMT, you have to be very careful with peak detection. Reporter ions do not have natural isotope distributions, so anything that assumes this will not be reliable.

Peak picking and search parameters

Mascot Distiller supports single peak picking in the reporter ion region



For TMTpro the labels make use of the mass defect between ^{13}C and ^{15}N

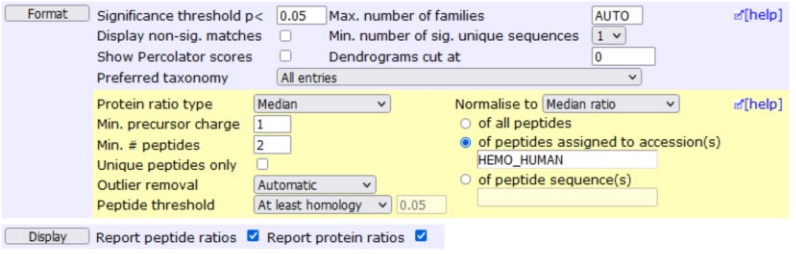


If you are using Mascot Distiller for the peak picking you can set the MS/MS Peak Picking parameters to treat the reporter ion region as single peaks with no deisotoping.

The TMTpro labels make use of the mass defect between ^{13}C and ^{15}N. This does not normally affect the peak picking, but we do need to use a narrower tolerance window in the reporter ion region in order to separate the peaks. This is specified in the quantitation method.

Whatever peak picking software you use, you'll probably need to experiment with the settings.

Normalisation

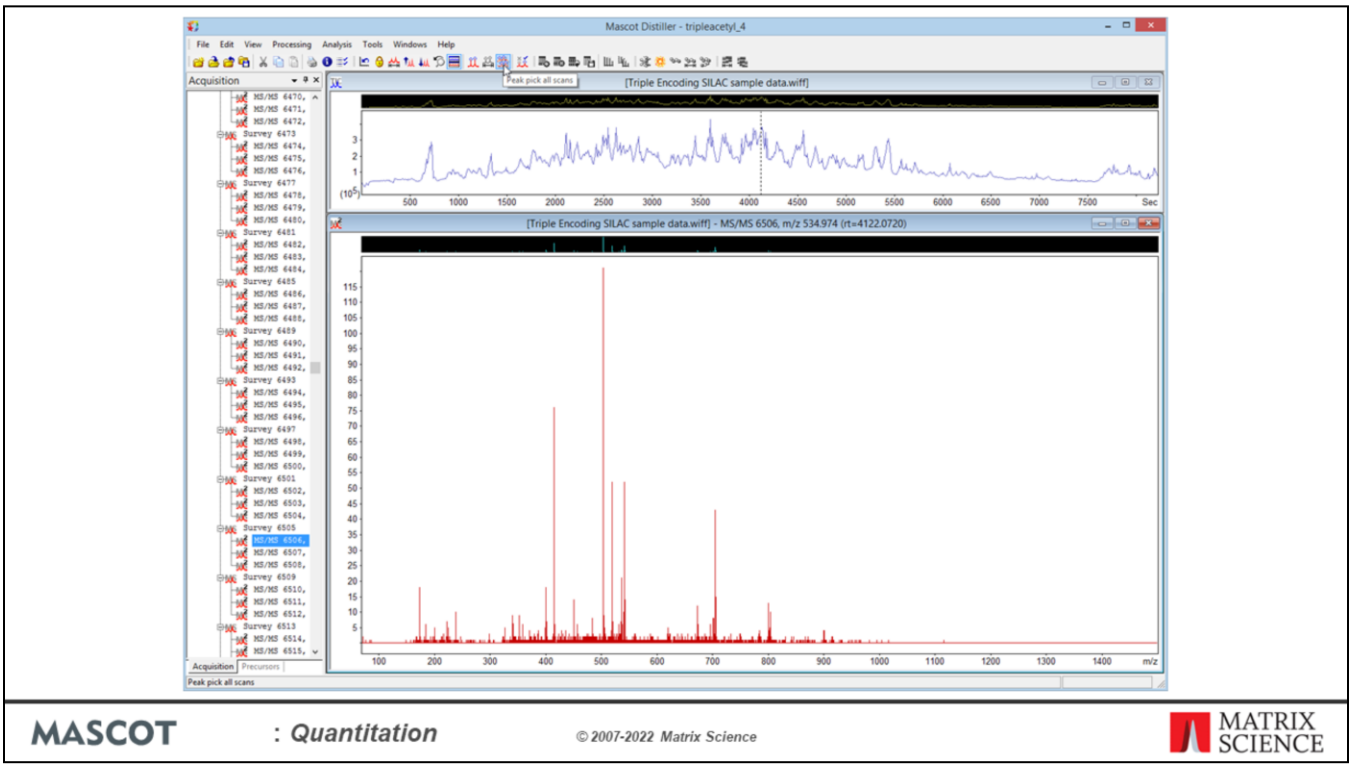


When to use global normalisation

- 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

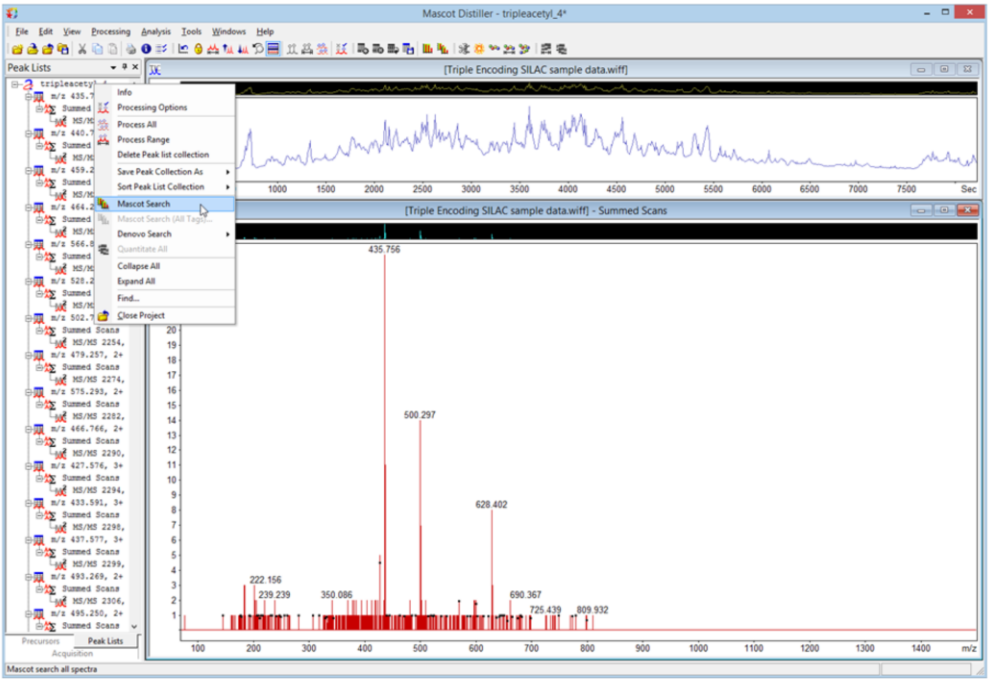
Whether to calculate protein ratios from the average, median or weighted average of the set of peptide ratios is best decided by running some standards (e.g. a cell lysate spiked with varying amounts of a known protein) and seeing which gives the best accuracy and precision.

Normalisation is a way to reduce or eliminate systematic errors. In Mascot 2.5, you can normalize 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 normalization 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

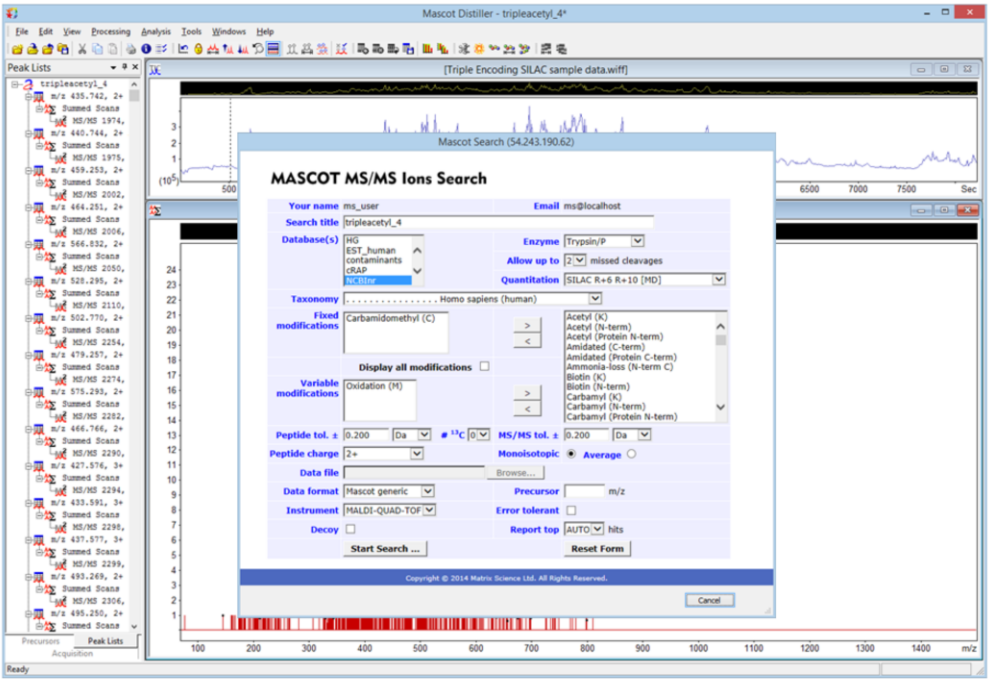


MASCOT : *Quantitation* MATRIX SCIENCE

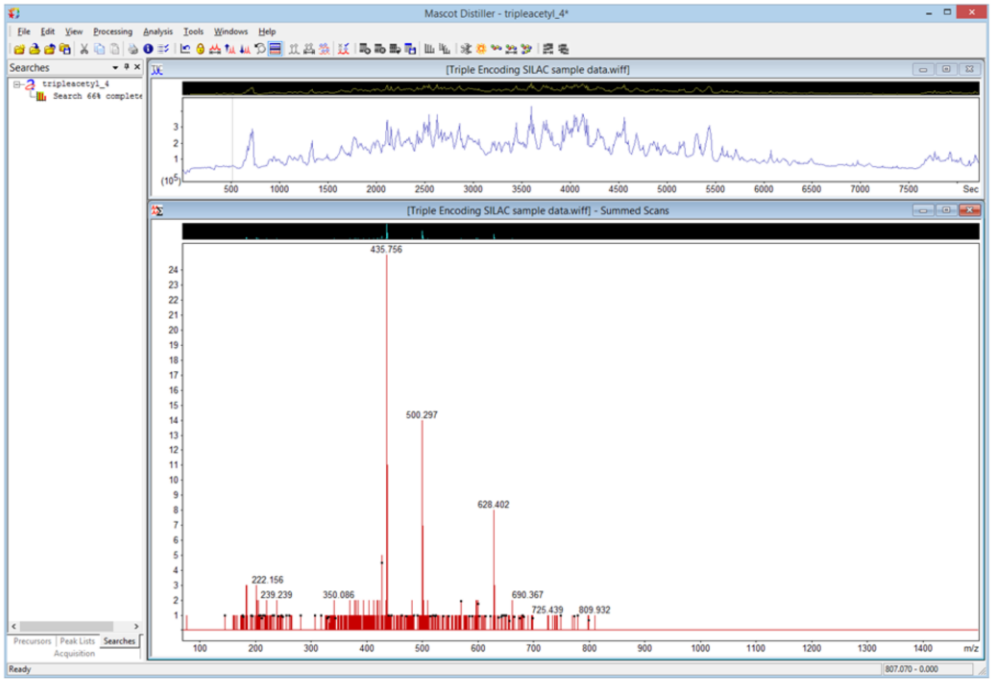
Lets turn to the precursor protocol. This is where we have to use the Mascot Distiller Quantitation Toolbox. I'm going to illustrate the workflow using one of the QStar sample data sets that can be downloaded from the MSQuant home page. This is a SILAC sample with three components: unlabelled, Arg labelled with 13C(6), and Arg labelled with 13C(6)15N(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

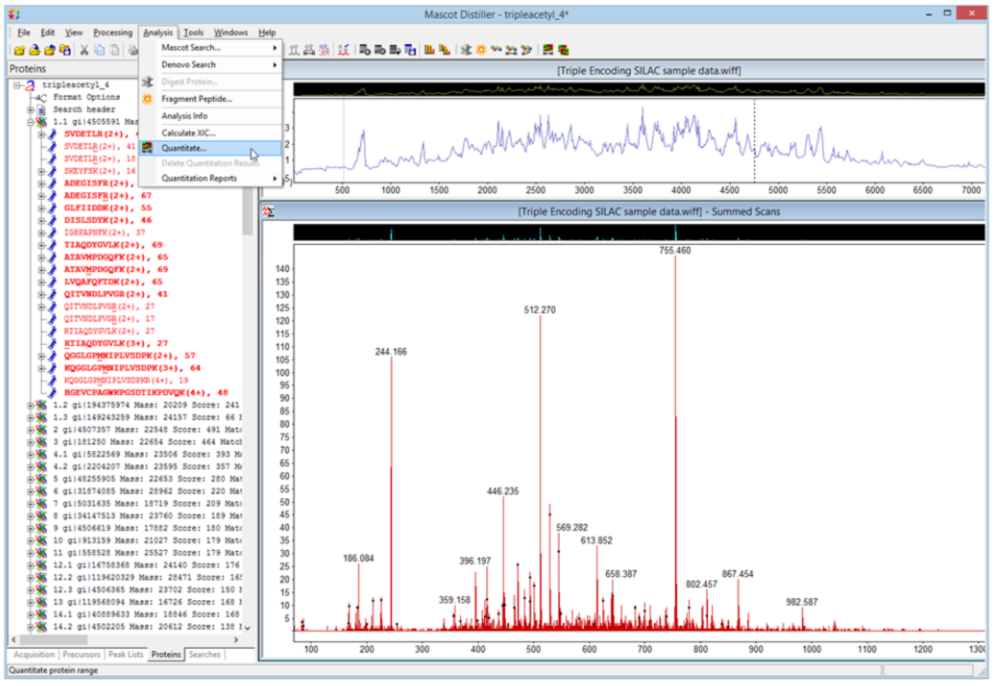


There's an opportunity to tweak the search conditions, but most of the important settings, including the modifications, are embedded in the selected quantitation method



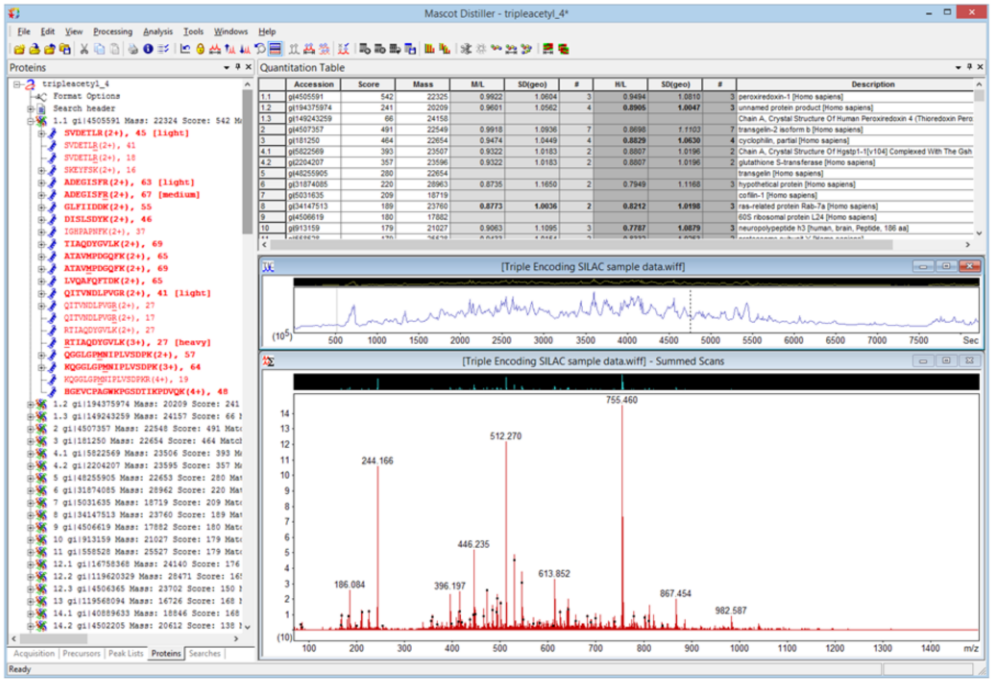
MASCOT : *Quantitation* MATRIX SCIENCE

We get progress reports while the search is running



MASCOT : *Quantitation* MATRIX SCIENCE

When the search is complete, the results are imported into Distiller. All the required information is now available, so Distiller is in a position to generate a quantitation report



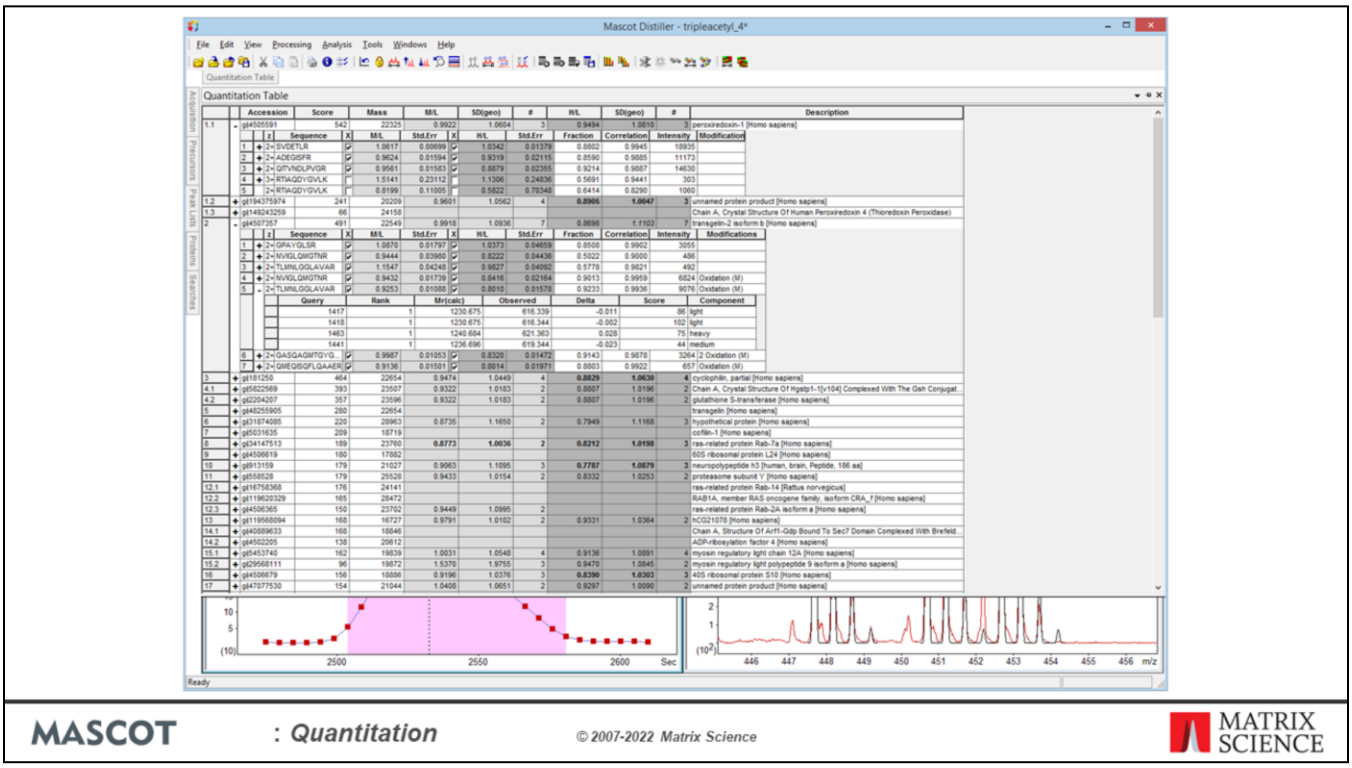
MASCOT : *Quantitation* MATRIX SCIENCE

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

MASCOT : *Quantitation* MATRIX SCIENCE

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.

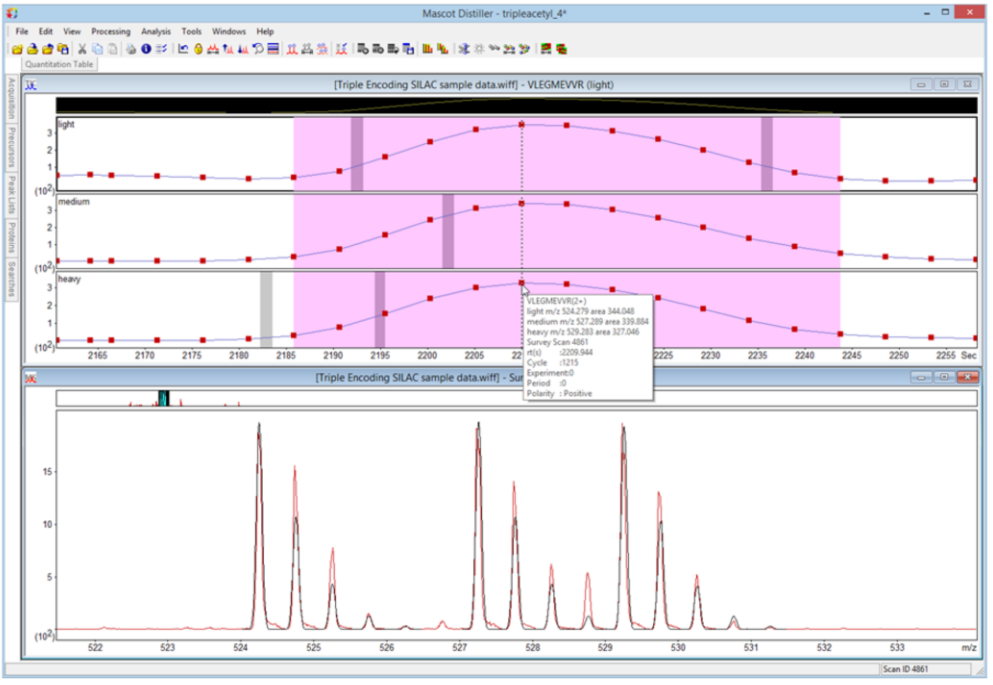
However big your screen, its always difficult to find enough room to display everything. To try and make best use of limited space, the grid and tree can be unpinned, so that they fly out when required and disappear when you move the mouse away.



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



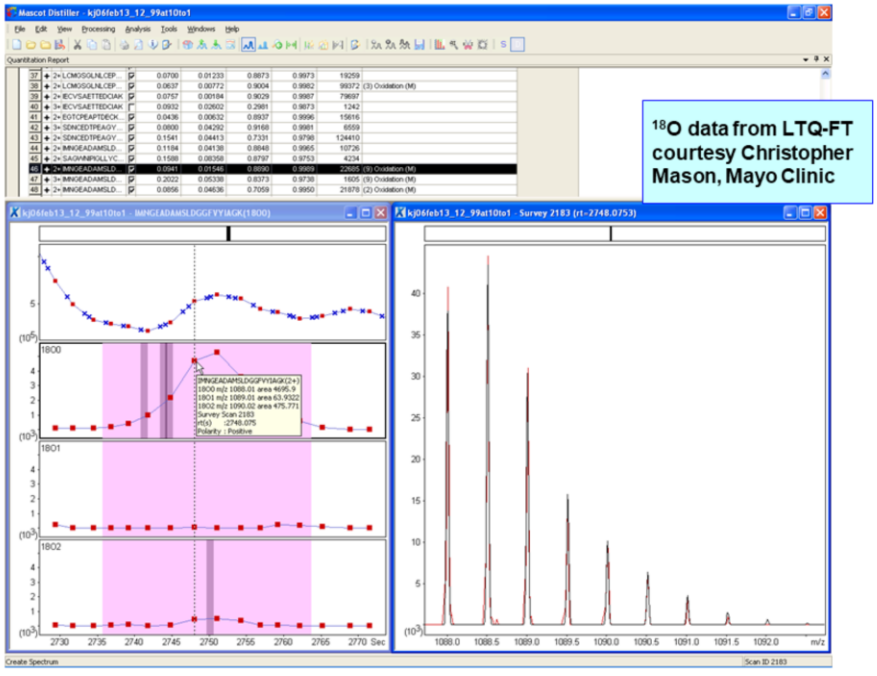
MASCOT : *Quantitation* MATRIX SCIENCE

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.

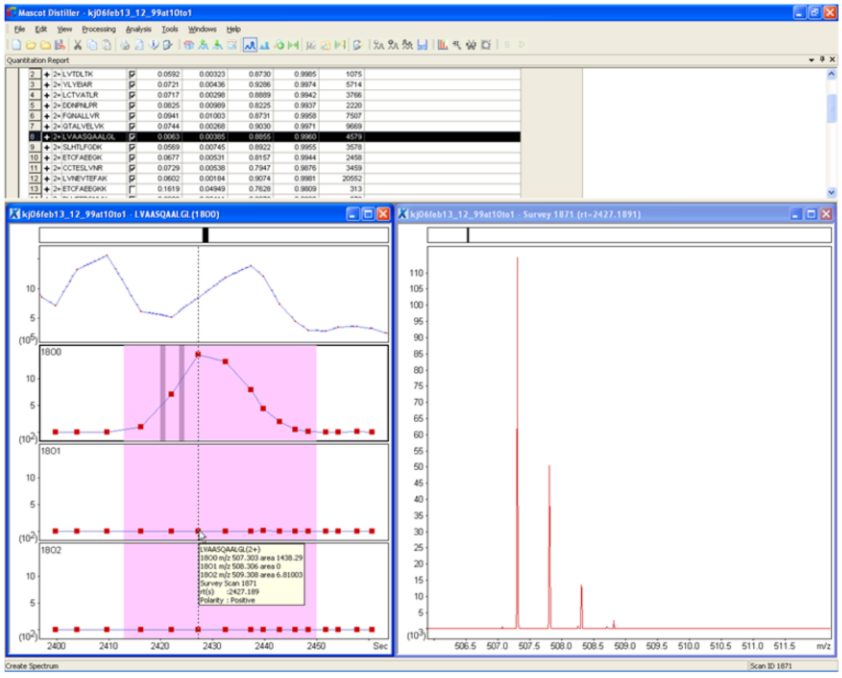
The light purple region is the XIC peak. These are the scans that have been integrated to determine the ratios. The grey bands, which look dark purple when within the XIC peak, are the scans for which we have Mascot matches. In this particular case, there are matches for all three components. If there 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



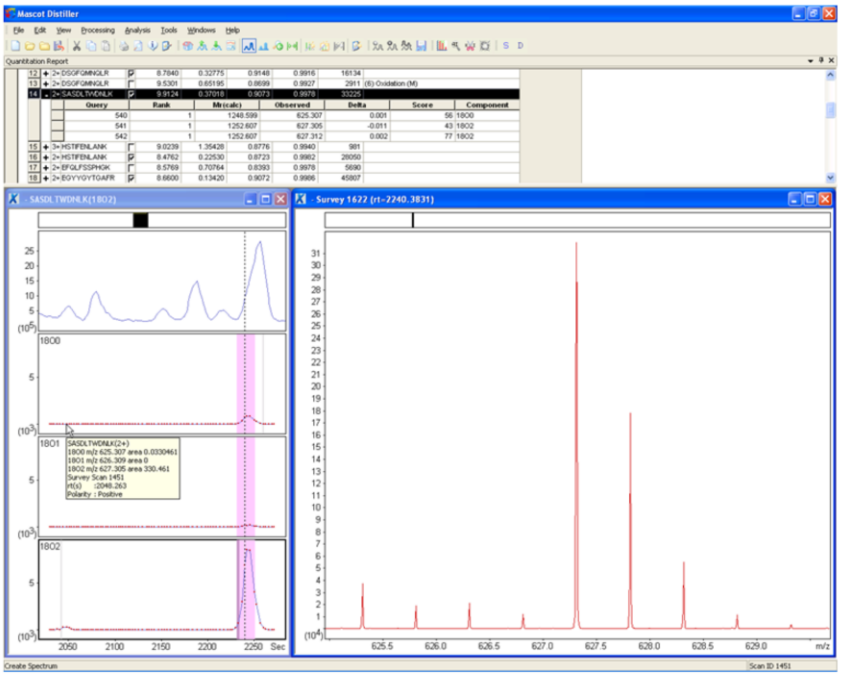
MASCOT : *Quantitation* MATRIX SCIENCE

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.



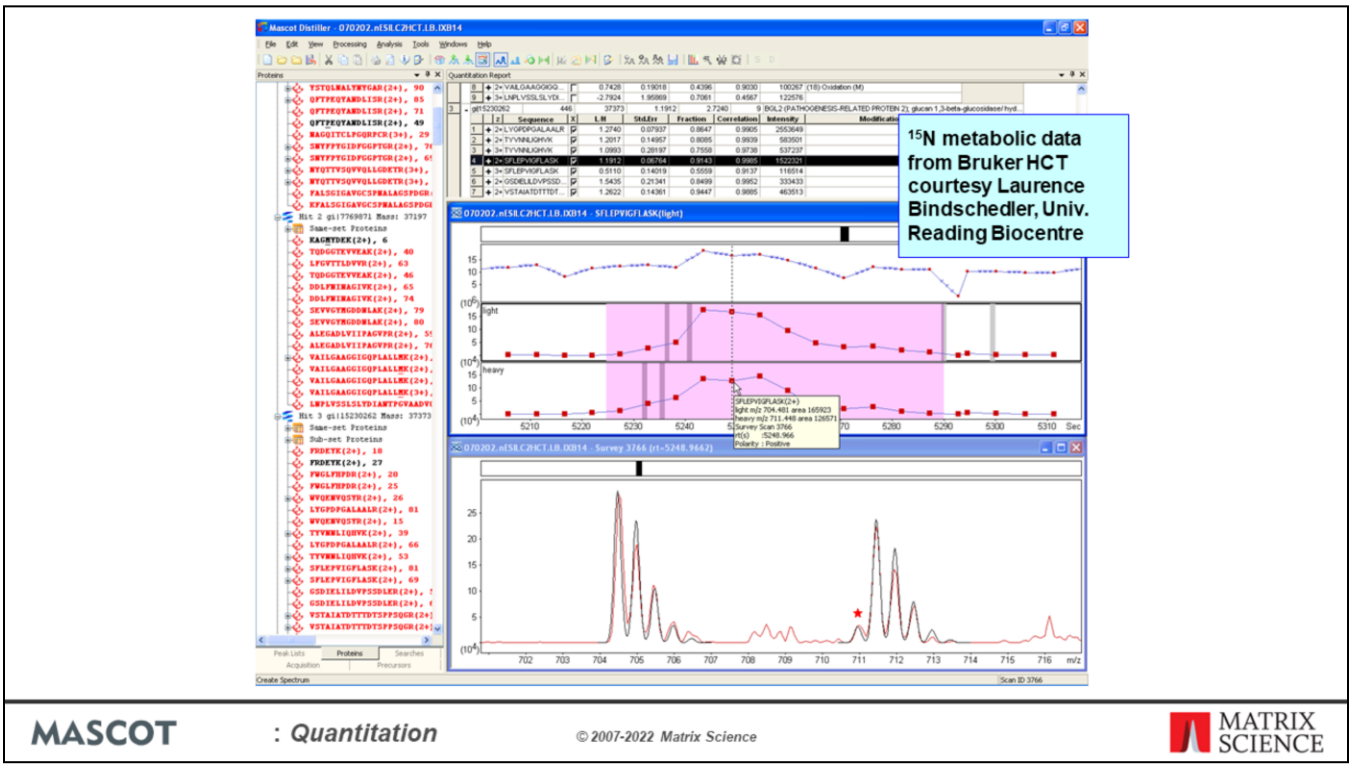
MASCOT : *Quantitation* MATRIX SCIENCE

This is the C-terminal peptide of BSA, which should be unlabelled, so reassuring to see the ratio is indeed close to zero



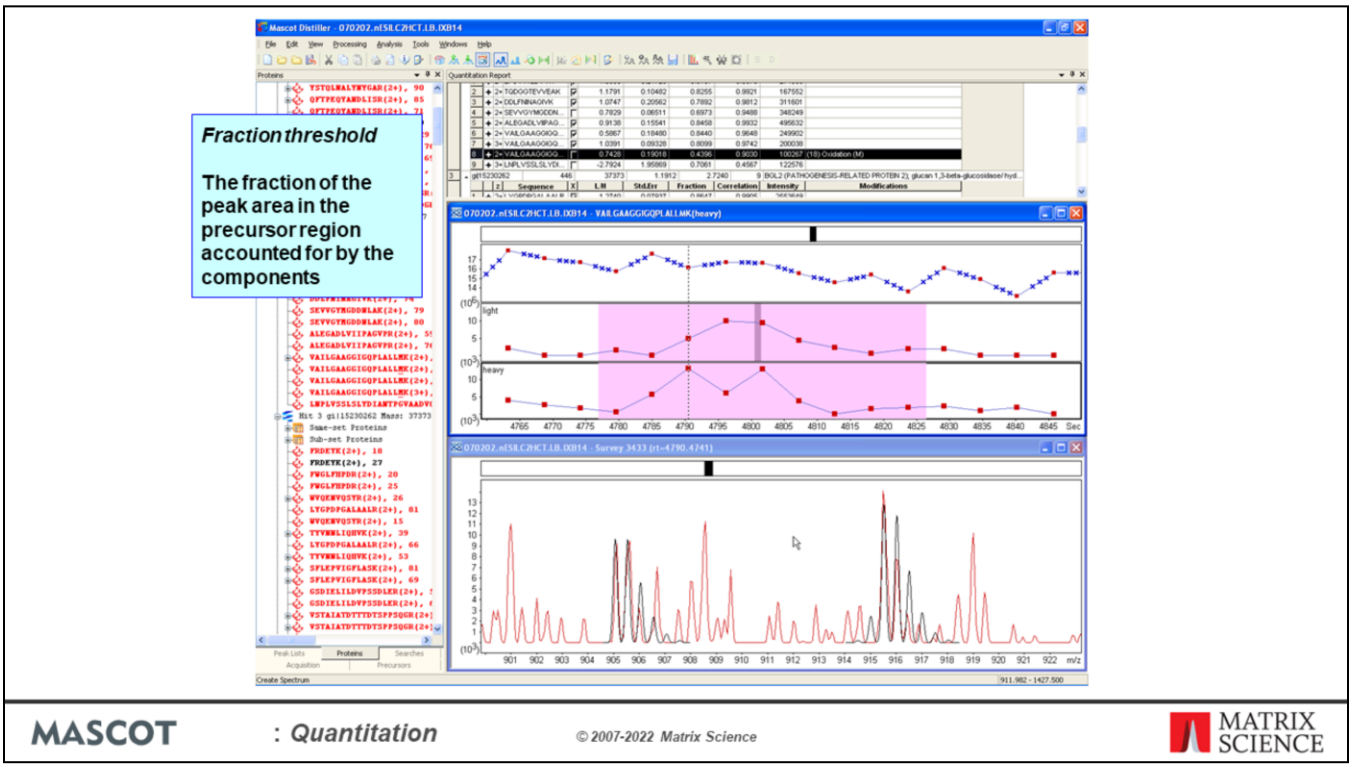
MASCOT : *Quantitation* MATRIX SCIENCE

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.

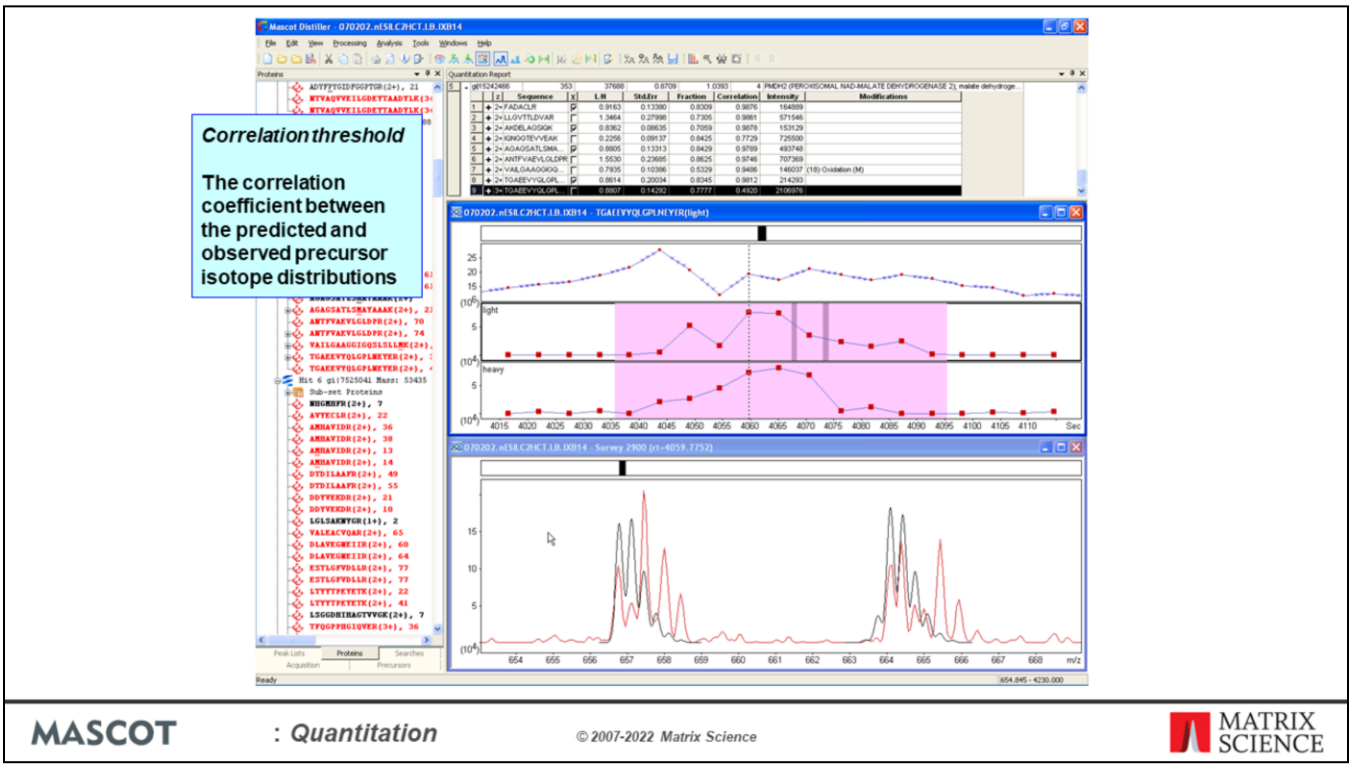


Now some metabolic data, courtesy of Rainer Cramer's group at the Reading Biocentre. In the heavy component, all the nitrogens in the proteins have been replaced by N-15. The calculated distributions (black) fit reasonably well to the observed peaks (red). Notice the peak marked with a red star. This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

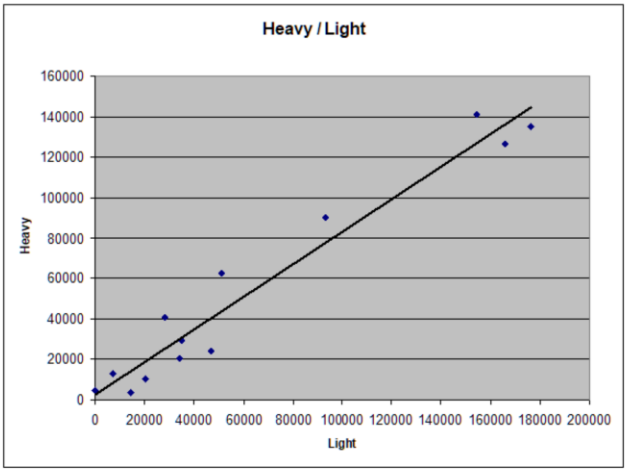


Fraction is the fraction of the peak area in the precursor region accounted for by the components. Here is a spectrum where there are a lot of interfering peaks. When we add up the areas, the expected precursors only account for 44% of the area, so the ratio is rejected. This threshold, like the others, 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 0.49

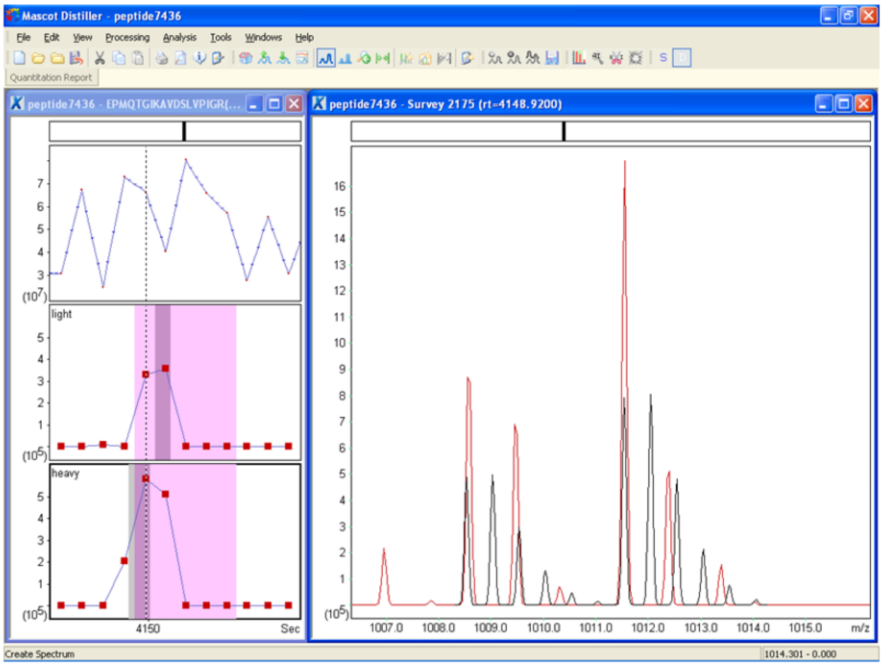
Quantitation - Precursor



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

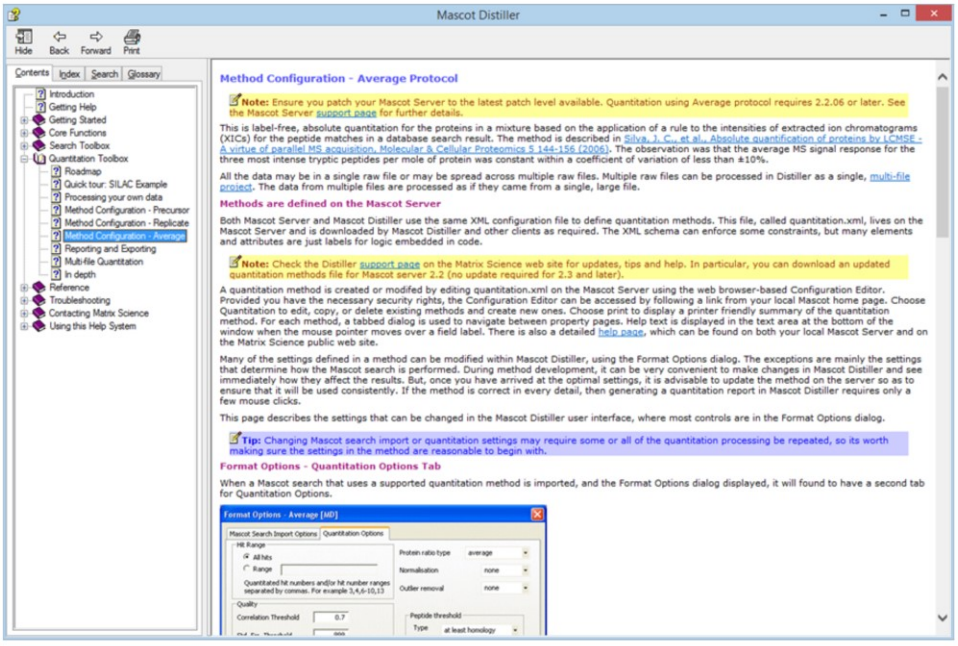
- The gradient is the best estimate of the ratio.
- Any background is corrected
- The standard error for the fit is a good measure of the reliability of the ratio

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.

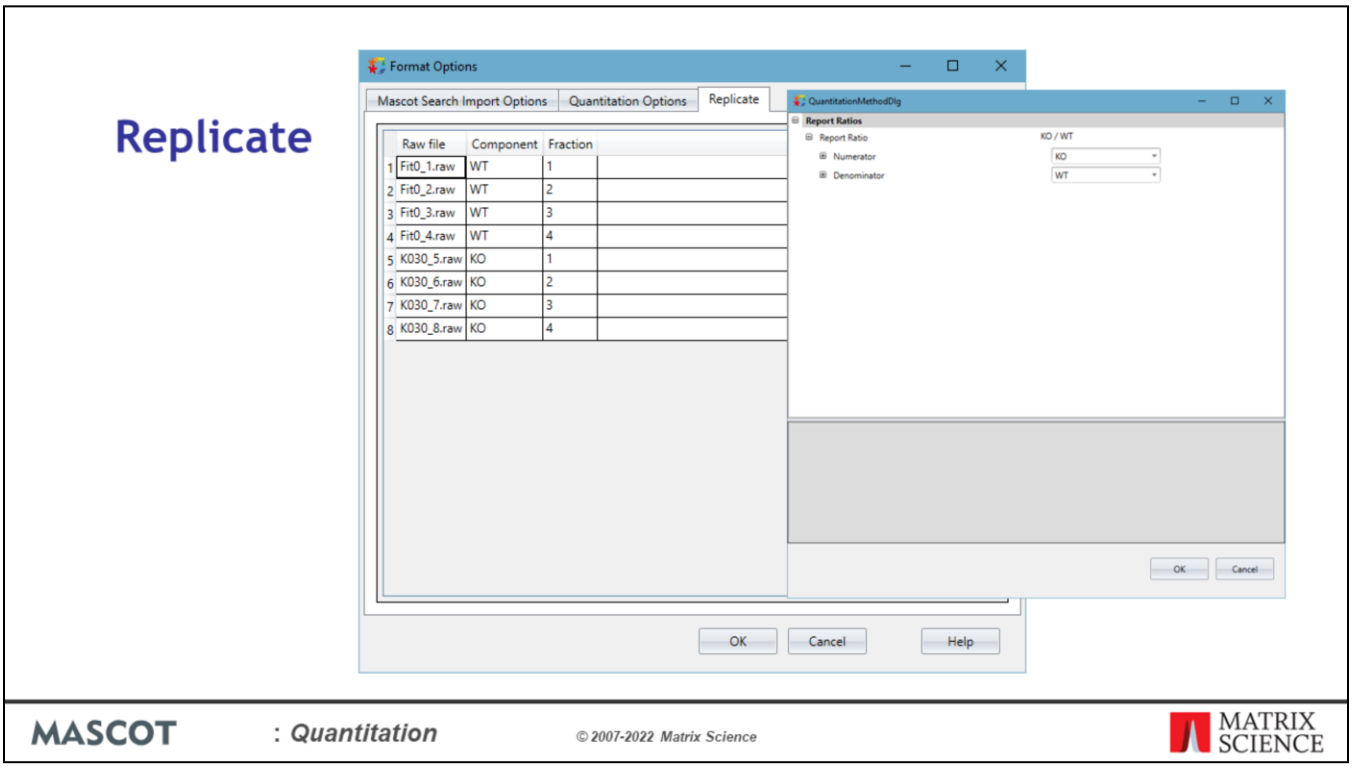


MASCOT : *Quantitation* MATRIX SCIENCE

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.

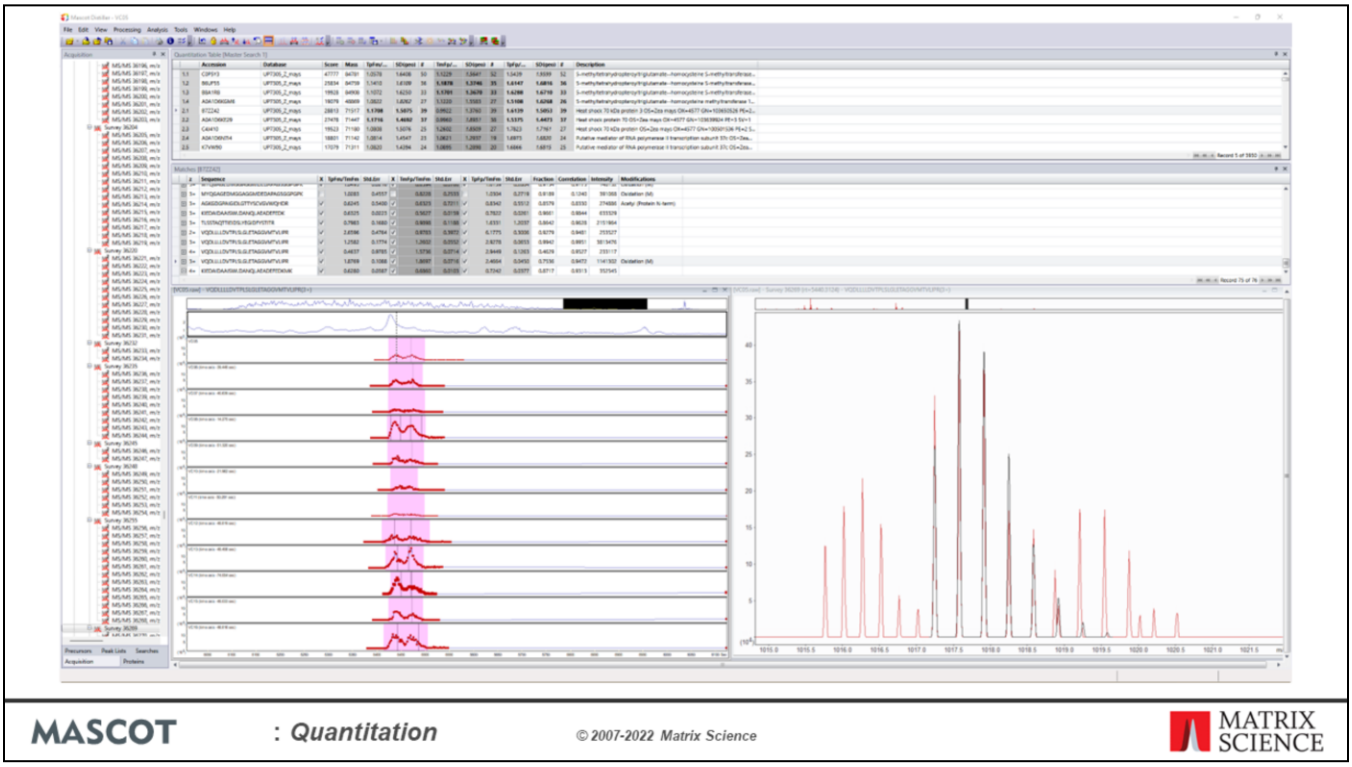


In the interests of time, we'll skip over the Average protocol, but use this opportunity to point out that Distiller comes with comprehensive help. This page gives an overview of how Average works and details of all the configuration settings. There are similar pages for precursor and replicate.

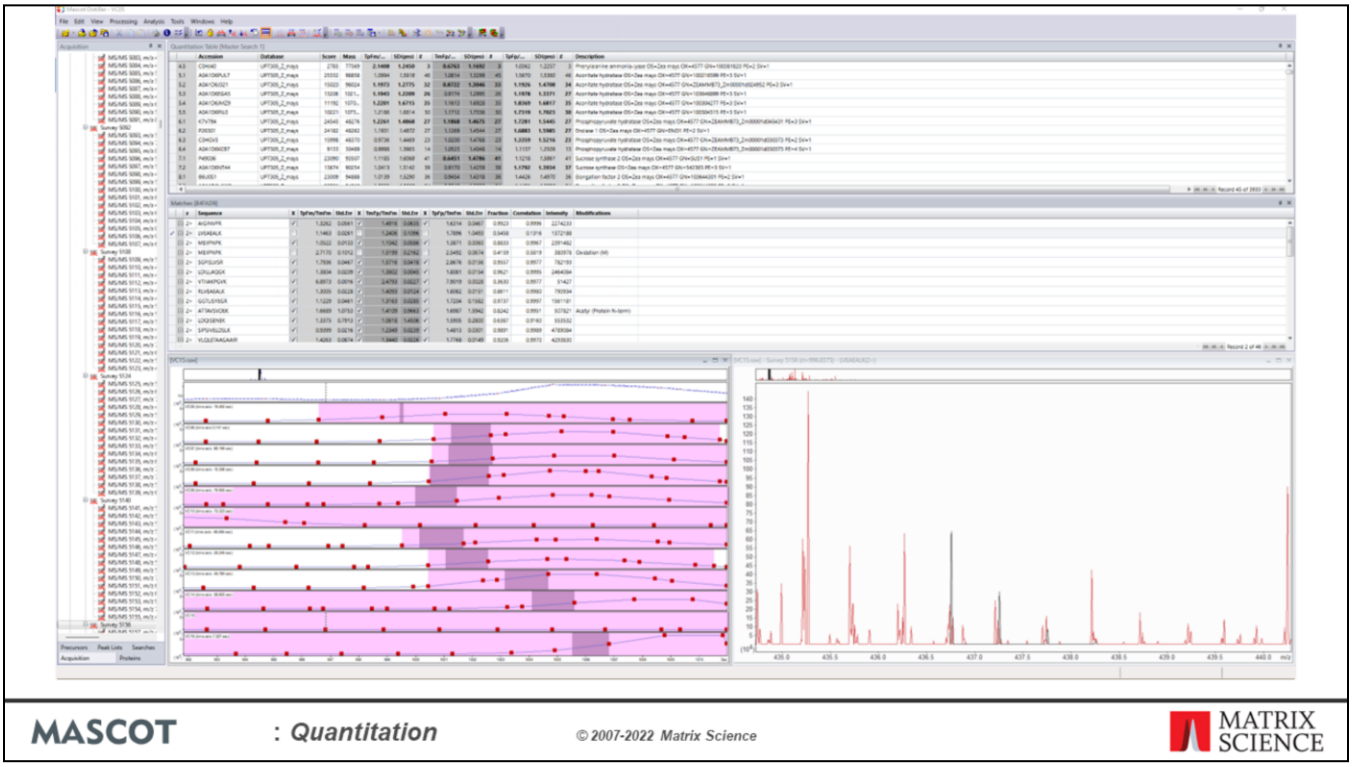


Replicate, you may remember, is ‘label free’. Our implementation is identification driven, not feature driven. Distiller starts from the MS/MS data, imports the peptide matches from a Mascot search, then aligns them against a consensus generated by roughly aligning and combining the total ion chromatograms (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 XIC’s are calculated for peptides identified in one file but the other(s). For fractionated samples, multiple consensuses 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.



The tricky part of label free is time alignment of the chromatograms. Distiller does its best, and here is an example where things are working, even though the precursor region is heavily overlapped. The Standard error, of the ratio measurement, fraction and correlation values measure the quality of the data and can be used to filter the peptides used for quantitation. Here, the fraction value is 0.75 which indicates most of the signal is coming from the query in question but proportion of it is not.



But, here it fails. Distiller can manage small misalignments OK, but not severe misalignments or multiple XIC peaks. If alignment fails for a particular peptide, then the ratio is meaningless. The MS1 chromatogram is from one of the runs, the first run by default, from a region where the peptide was successfully identified and as such it represents one of the better views of the precursor. The fraction value is reported as 0.54 with a peak correlation of 0.136. Both these values indicate the data is not going to be reliable for this peptide. Clicking on a different fraction and location we can see why these values are quite low.

Reporting

Basic tables of peptide and protein information
Quality control report
XML file

Quantitation summaries via Mascot Daemon

The Mascot Distiller GUI is a good way to review the quantitation results, to see that everything has worked as expected.

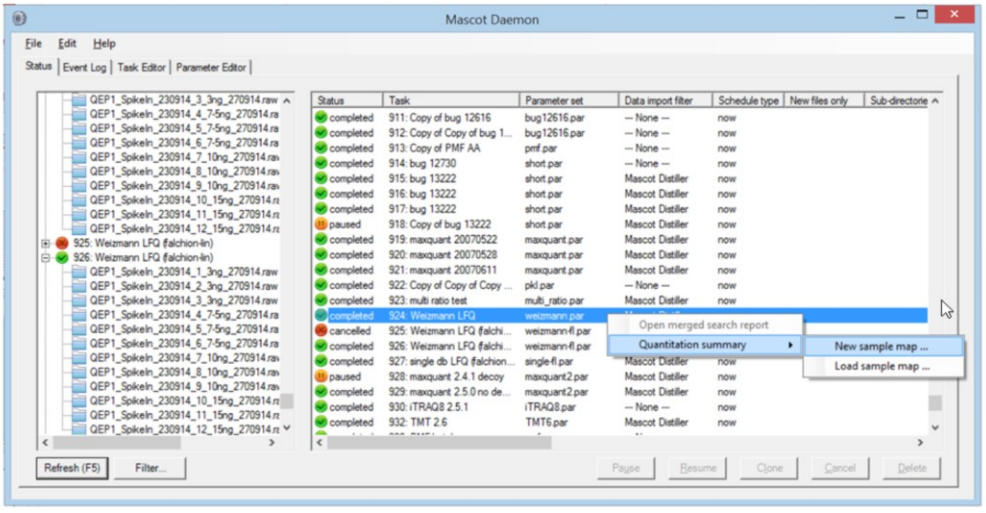
The raw results can be used to generate protein and peptide tables in a html report format or excel files that can be used for further analysis. The table+peptides_int table presents the results in a format ready for peptide level quantitation that is useful when running phosphopeptide quantitation experiments.

There is a simple quality control report.

And an XML format that can be used by third party software. Proteome Solutions Scaffold Q+S uses this format

New in Mascot Daemon 2.7 is the feature to export results from a Mascot Daemon task into a quantitation summary.

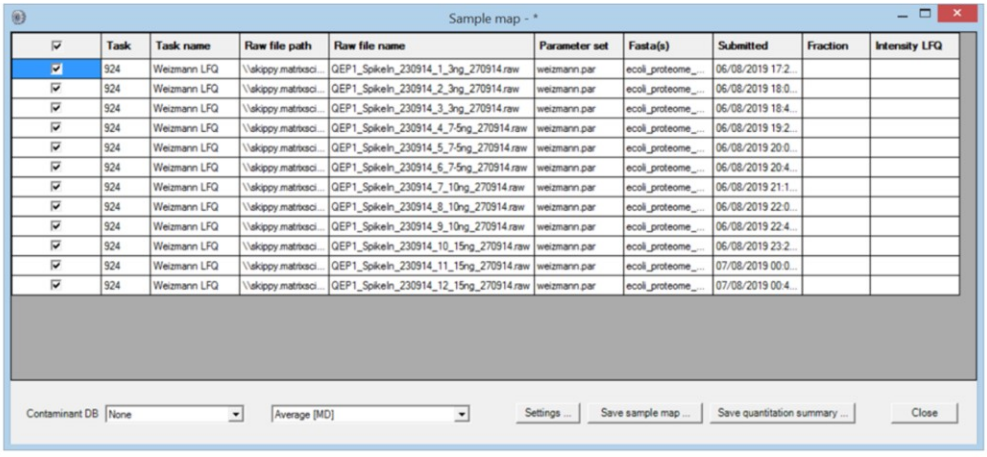
Quantitation summary



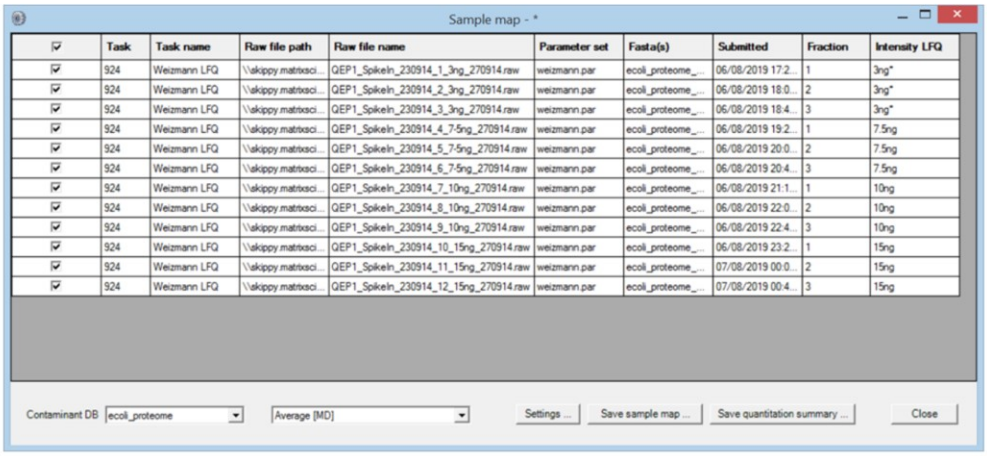
The Quantitation summary is a method to output the quantitation results in a format ready to use for statistical analysis. You can export the results from any quantitation method where you have used Mascot Daemon to automate the processing.

Here the files were processed in a single Mascot Daemon task, using Mascot Distiller for peak picking and quantitation. The data is a Label Free Quantitation (LFQ) data set and searched with the 'Average [MD]' quantitation method.

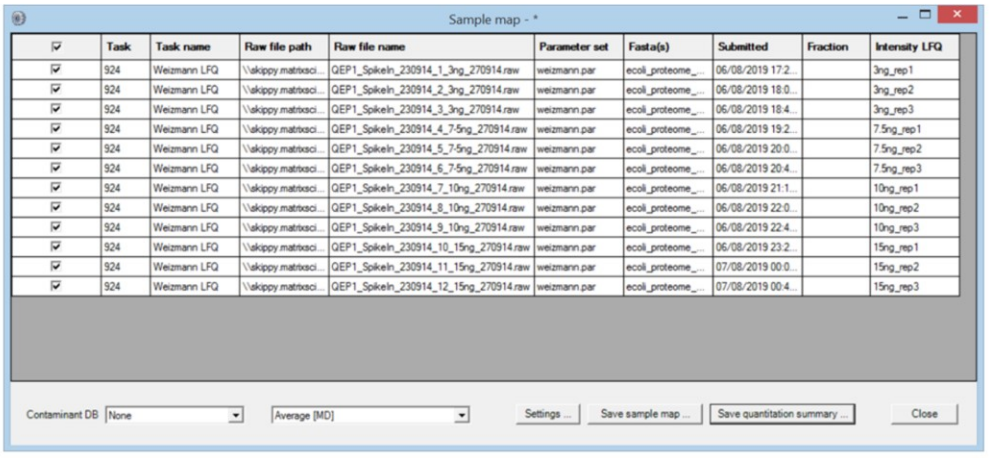
Once processing was complete, the task was selected in the list view on the Mascot Daemon status tab. Right clicking the selection invoked a context menu, from which Quantitation Summary; New sample map ... was chosen.



We've tried to minimise the amount of typing required to create a Sample Map. As in Excel, columns can be sorted so that a repeating value can be pasted to a range of cells. In this case, sorting on raw file name (by clicking the column header) is all that is required. For more complex data sets, sorting on file path or task name or time of submission may help organise the files in a useful way. This is particularly important when samples have been separated into large numbers of fractions. You don't want to have to type in every fraction number. Just sort appropriately, select the cell range in the fraction column, right click and choose Fill with integer series.

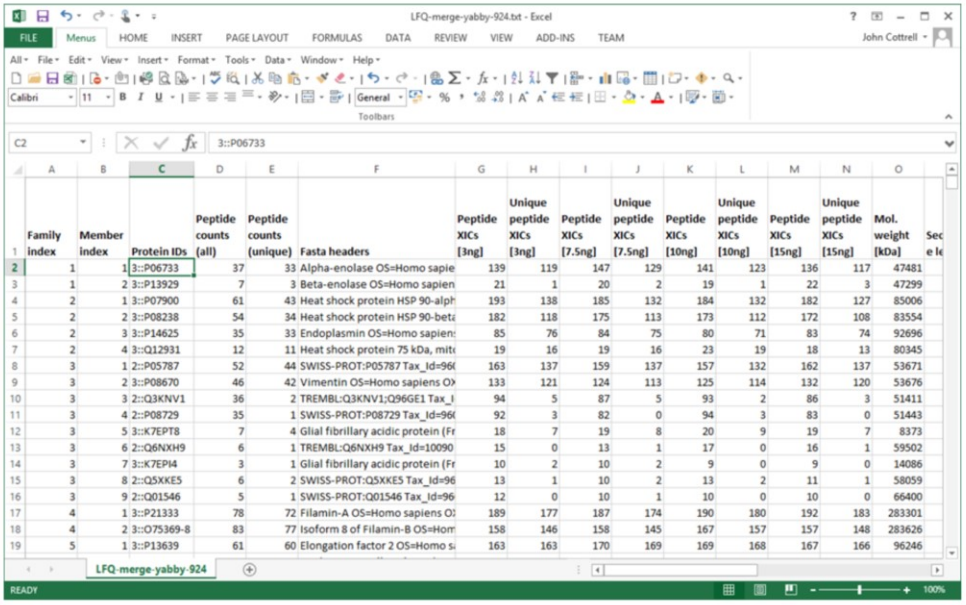


If a sample is not fractionated, the fraction cells can be left empty, unless you wish to merge replicates by treating them as fractions. In this example, there are two useful ways to fill in the two columns on the right that are used to identify the samples. Like this, to merge replicates as if they were fractions. An asterisk indicates the reference or control sample, and the Quantitation Summary will include ratios relative to this sample.



Or, like this, to create separate columns in the Quantitation Summary for each replicate; useful if you want statistics for variation across replicates. Sample identifiers can be anything you like as long as the combination of identifier and fraction number for each file is unique.

The Sample Map can be saved to a disk file, even if not complete, and reloaded as required. When Save quantitation summary ... is chosen, some validation is performed.

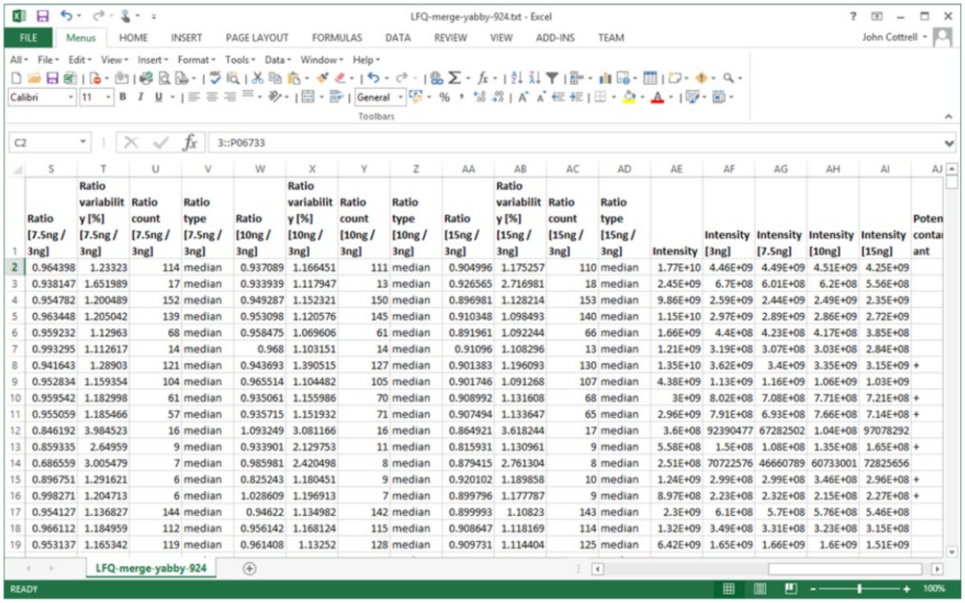


	Family index	Member index	Protein IDs	Peptide counts (all)	Peptide counts (unique)	Fasta headers	Peptide XICs [3ng]	Unique peptide XICs [3ng]	Peptide XICs [7.5ng]	Unique peptide XICs [7.5ng]	Peptide XICs [10ng]	Unique peptide XICs [10ng]	Peptide XICs [15ng]	Unique peptide XICs [15ng]	Mol. weight [kDa]
2	1	1	3::P06733	37	33	Alpha-enolase OS=Homo sapie	139	119	147	129	141	123	136	117	47481
3	1	2	3::P13929	7	3	Beta-enolase OS=Homo sapien	21	1	20	2	19	1	22	3	47299
4	2	1	3::P07900	61	43	Heat shock protein HSP 90-alph	193	138	185	132	184	132	182	127	85006
5	2	2	3::P08238	54	34	Heat shock protein HSP 90-beta	182	118	175	113	173	112	172	108	83554
6	2	3	3::P14625	35	33	Endoplasmin OS=Homo sapiens	85	76	84	75	80	71	83	74	92696
7	2	4	3::Q12931	12	11	Heat shock protein 75 kDa, mit	19	16	19	16	23	19	18	13	80345
8	3	1	2::P05787	52	44	SWISS-PROT:P05787 Tax_Id=96	163	137	159	137	157	132	162	137	53671
9	3	2	3::P08670	46	42	Vimentin OS=Homo sapiens OX	133	121	124	113	125	114	132	120	53676
10	3	3	2::Q3KNV1	36	2	TREMBL:Q3KNV1;Q96GE1 Tax_I	94	5	87	5	93	2	86	3	51411
11	3	4	2::P08729	35	1	SWISS-PROT:P08729 Tax_Id=96	92	3	82	0	94	3	83	0	51443
12	3	5	3::K7EPT8	7	4	Glial fibrillary acidic protein (Fr	18	7	19	8	20	9	19	7	8373
13	3	6	2::Q6NXH9	6	1	TREMBL:Q6NXH9 Tax_Id=10090	15	0	13	1	17	0	16	1	59502
14	3	7	3::K7EPI4	3	1	Glial fibrillary acidic protein (Fr	10	2	10	2	9	0	9	0	14086
15	3	8	2::Q5XKE5	6	2	SWISS-PROT:Q5XKE5 Tax_Id=96	13	1	10	2	13	2	11	1	58059
16	3	9	2::Q01546	5	1	SWISS-PROT:Q01546 Tax_Id=96	12	0	10	1	10	0	10	0	66400
17	4	1	3::P21333	78	72	Filamin-A OS=Homo sapiens O	189	177	187	174	190	180	192	183	283301
18	4	2	3::O75369-8	83	77	Isoform 8 of Filamin-B OS=Hom	158	146	158	145	167	157	157	148	283626
19	5	1	3::P13639	61	60	Elongation factor 2 OS=Homo s	163	163	170	169	169	168	167	166	96246

If all is present and correct, a progress dialog is displayed, because calculations can take some time for large numbers of files. The stages are

- Create a merged report of all search results
- Export the master list of proteins
- For each file, export the expression data for all peptides
- Assign the peptide data to proteins according to the master list
- Calculate protein abundances and ratios as required, including outlier detection
- Write everything to a disk file in TSV format

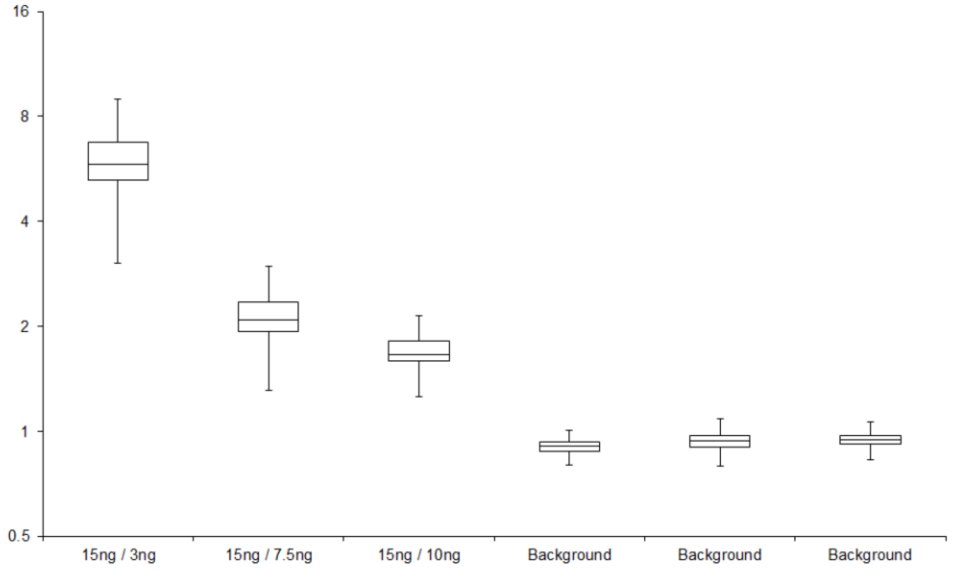
This is the Quantitation Summary for the label-free data when we choose to merge replicates. Most columns are self-explanatory



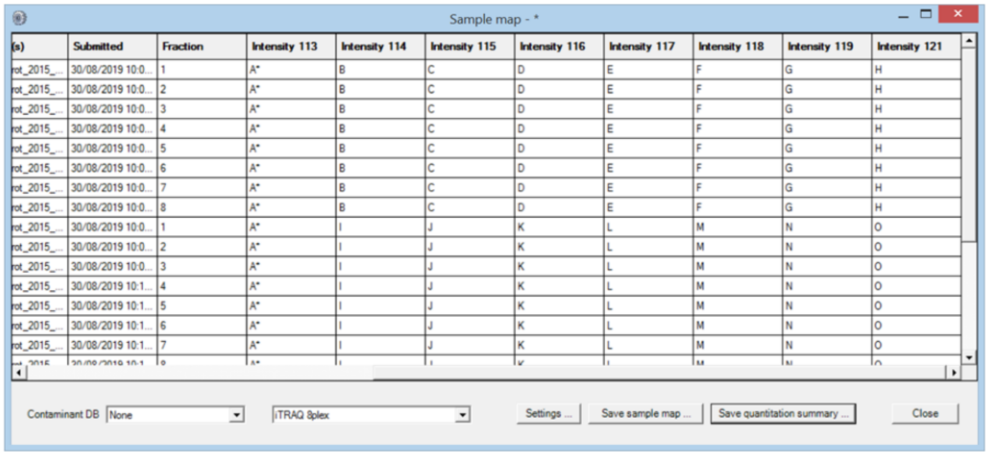
LFQ-merge-yabby-924.txt - Excel

	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ
1	Ratio [7.5ng / 3ng]	Ratio variability [%] [7.5ng / 3ng]	Ratio count [7.5ng / 3ng]	Ratio type [7.5ng / 3ng]	Ratio [10ng / 3ng]	Ratio variability [%] [10ng / 3ng]	Ratio count [10ng / 3ng]	Ratio type [10ng / 3ng]	Ratio [15ng / 3ng]	Ratio variability [%] [15ng / 3ng]	Ratio count [15ng / 3ng]	Ratio type [15ng / 3ng]	Intensity	Intensity [3ng]	Intensity [7.5ng]	Intensity [10ng]	Intensity [15ng]	Potential contaminant
2	0.964398	1.23323	114	median	0.937089	1.166451	111	median	0.904996	1.175257	110	median	1.77E+10	4.46E+09	4.49E+09	4.51E+09	4.25E+09	
3	0.938147	1.651989	17	median	0.933939	1.117947	13	median	0.926565	2.716981	18	median	2.45E+09	6.7E+08	6.01E+08	6.2E+08	5.56E+08	
4	0.954782	1.200489	152	median	0.949287	1.152321	150	median	0.896981	1.128214	153	median	9.86E+09	2.59E+09	2.44E+09	2.49E+09	2.35E+09	
5	0.963448	1.205042	139	median	0.953098	1.120576	145	median	0.910348	1.098493	140	median	1.15E+10	2.97E+09	2.89E+09	2.86E+09	2.72E+09	
6	0.959232	1.12963	68	median	0.958475	1.069606	61	median	0.891961	1.092244	66	median	1.66E+09	4.4E+08	4.23E+08	4.17E+08	3.85E+08	
7	0.993295	1.112617	14	median	0.968	1.103151	14	median	0.91096	1.108296	13	median	1.21E+09	3.19E+08	3.07E+08	3.03E+08	2.84E+08	
8	0.941643	1.28903	121	median	0.943693	1.390515	127	median	0.901383	1.196093	130	median	1.35E+10	3.62E+09	3.4E+09	3.35E+09	3.15E+09	+
9	0.952834	1.159354	104	median	0.965514	1.104482	105	median	0.901746	1.091268	107	median	4.38E+09	1.13E+09	1.16E+09	1.06E+09	1.03E+09	
10	0.959542	1.182998	61	median	0.935061	1.155986	70	median	0.908992	1.131608	68	median	3E+09	8.02E+08	7.08E+08	7.71E+08	7.21E+08	+
11	0.955059	1.185466	57	median	0.935715	1.151932	71	median	0.907494	1.133647	65	median	2.96E+09	7.91E+08	6.93E+08	7.66E+08	7.14E+08	+
12	0.846192	3.984523	16	median	1.093249	3.081166	16	median	0.864921	3.618244	17	median	3.6E+08	92390477	67282502	1.04E+08	97078292	
13	0.859335	2.64959	9	median	0.933901	2.129753	11	median	0.815931	1.130961	9	median	5.58E+08	1.5E+08	1.08E+08	1.35E+08	1.65E+08	+
14	0.686559	3.005479	7	median	0.985981	2.420498	8	median	0.879415	2.761304	8	median	2.51E+08	70722576	46660789	60733001	72825656	
15	0.896751	1.291621	6	median	0.825243	1.180451	9	median	0.920102	1.189858	10	median	1.24E+09	2.99E+08	2.99E+08	3.46E+08	2.96E+08	+
16	0.998271	1.204713	6	median	1.028609	1.196913	7	median	0.899796	1.177787	9	median	8.97E+08	2.23E+08	2.32E+08	2.15E+08	2.27E+08	+
17	0.954127	1.136827	144	median	0.94622	1.134982	142	median	0.899993	1.10823	143	median	2.3E+09	6.1E+08	5.7E+08	5.76E+08	5.46E+08	
18	0.966112	1.184959	112	median	0.956142	1.168124	115	median	0.908647	1.118169	114	median	1.32E+09	3.49E+08	3.31E+08	3.23E+08	3.15E+08	
19	0.953137	1.165342	119	median	0.961408	1.13252	128	median	0.909731	1.114404	125	median	6.42E+09	1.65E+09	1.66E+09	1.6E+09	1.51E+09	

Over to the right are columns containing median ratios and total intensity values for each protein. One of the strengths of the Quantitation Summary is that it uses the same rigorous protein inference as the Mascot Protein Family Summary report.



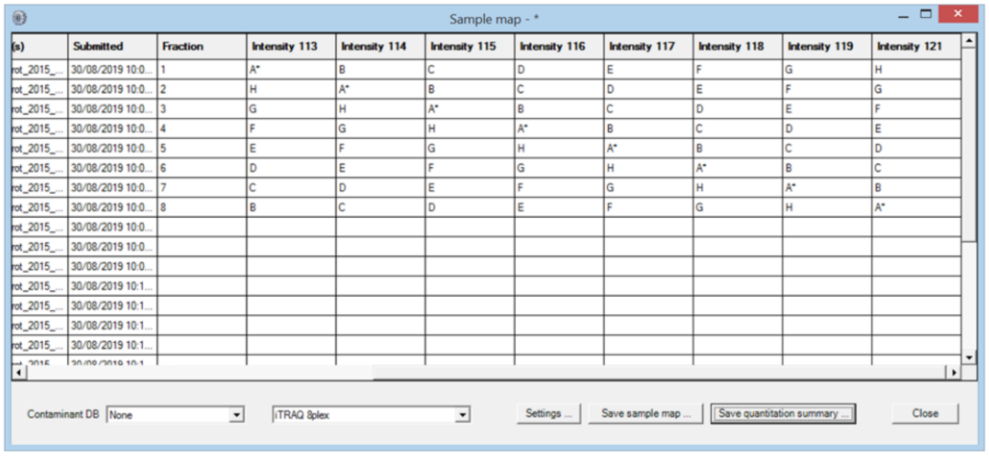
You may be able to get the report you want direct from Excel. For example, this box and whisker plot of the data was produced in Excel.



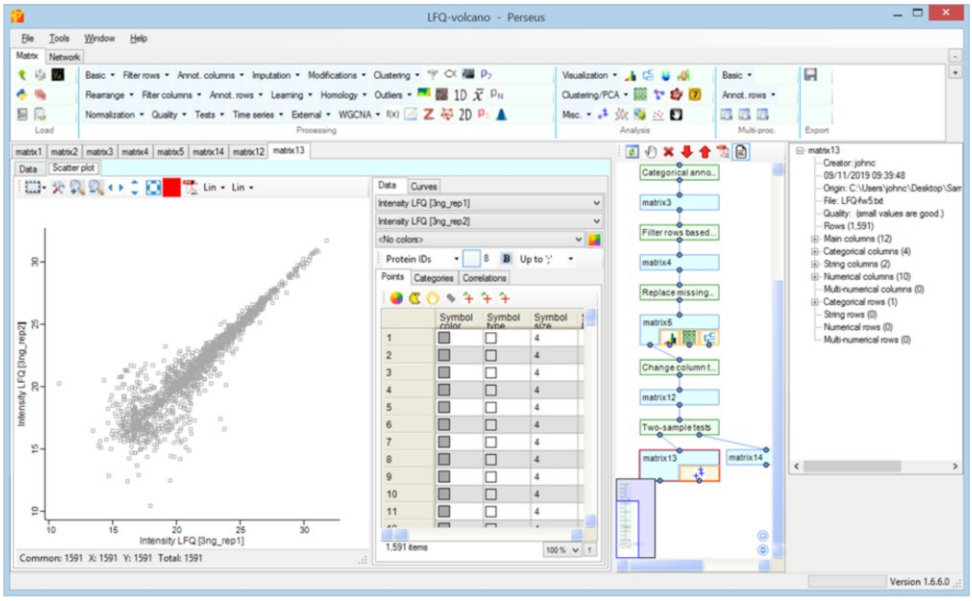
For a label-free experiment, there is a single column for the sample identifier. For experiments that use isotopic labels, there will be a column for each component specified in the quantitation method. If it was a typical SILAC experiment with two components, light for unlabelled and heavy for labelled, there would be two columns labelled *Intensity light* and *Intensity heavy*. An experiment that uses isobaric tags might have eight or more components.

This is a sample map for 8plex iTRAQ data. There are many ways of conducting such an study. This shows a case where there are 8 fractions for each sample, so the first 8 rows shows the same arrangement of samples, A to H. These fractions will be merged in the Quantitation summary, and A has an asterisk, so there will be columns for ratios to sample A as well as the total intensities for each channel. The second set of rows contains 7 new samples, plus reference sample A.

If the rows were replicates, and not fractions, then using the same channel for a sample across multiple replicates would be missing a trick.



Better to rotate the labels, so as to reduce or eliminate systematic errors. Ideally, a so-called Latin Square, where each sample is rotated through all possible tags, as shown here for the first 8 rows. Rows are merged by sample identifier, so that the Quantitation Summary contains the correct ratio and intensity information.



Having created a Quantitation Summary, what can you do with it? One option is to open it in Perseus, from the Max Planck Institute. This is a good choice if you prefer to manipulate the data using a spreadsheet type of approach. If you are willing to do a bit of scripting, the R language provides access to a huge range of statistical and graphical tools.

Miscellaneous

- Arg-Pro Conversion of SILAC label is supported
- Quantitation can be limited to 'unique' peptides
- Quantitation results, including all intermediate values, can be exported as XML and CSV
- Reporting uses Python 3.x scripts.
- Editable and extendable by Python programmers
- Exporting to CSV will open the report in Excel
- All steps from peak picking to quantitation can be automated using Mascot Daemon

Some final points

Selected Literature

- Ross, P. L., et al., *Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents*, Molecular & Cellular Proteomics 3 1154-1169 (2004) - iTRAQ
- Zhang, G. A. and Neubert, T. A., *Automated comparative proteomics based on multiplex tandem mass spectrometry and stable isotope labeling*, Molecular & Cellular Proteomics 5 401-411 (2006) - Multiplex
- Beynon, R. J. and Pratt, J. M., *Metabolic labeling of proteins for proteomics*, Molecular & Cellular Proteomics 4 857-872 (2005) - Metabolic
- Ong, S. E. and Mann, M., *Mass spectrometry-based proteomics turns quantitative*, Nature Chemical Biology 1 252-262 (2005) - General review
- Lill, J., *Proteomic tools for quantitation by mass spectrometry*, Mass Spectrometry Reviews 22 182-194 (2003) - General review
- Julka, S. and Regnier, F., *Quantification in proteomics through stable isotope coding: A review*, Journal of Proteome Research 3 350-363 (2004) - General review
- Bantscheff, M., et al., Quantitative mass spectrometry in proteomics: a critical review, Analytical and Bioanalytical Chemistry 389 1017-1031 (2007) - General review