

Quantitation

MASCOT



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
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

MASCOT : Quantitation

© 2007-2014 Matrix Science



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)



MASCOT : Quantitation

© 2007-2014 Matrix Science

MATRIX
SCIENCE

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

The screenshot shows the MASCOT MS/MS Ions Search web interface. The browser address bar displays the URL: 54.243.190.62/mascot/cgi/search_form.pl?FORMVER=2&SEARCH=MIS. The page title is "MASCOT MS/MS Ions Search".

Key fields and options visible include:

- Your name: ms_user, Email: msdlocalhost
- Search title: (empty)
- Databases(s): CRAP, NCBI, SwissProt, Uniprot_human, Uniprot_mouse
- Enzyme: TrypsinP
- Allow up to: 2 missed cleavages
- Quantitation: (ITRAQ 4plex) - This dropdown menu is open, showing a list of methods: mTRAQ (N-term), mTRAQ (Y), mTRAQ 13C(3)15N(1) (K), mTRAQ 13C(3)15N(1) (N-term), mTRAQ 13C(3)15N(1) (Y), mTRAQ 13C(6)15N(2) (K), mTRAQ 13C(6)15N(2) (N-term), mTRAQ 13C(6)15N(2) (Y), NIPCAM (C), Oxidation (PW), Phospho (ST).
- Taxonomy: All entries
- Fixed modifications: Carbamidomethyl (C)
- Variable modifications: Oxidation (M)
- Peptide tol.: 20 ppm, # 13C: 0, MS/MS tol.: 0.1 Da
- Peptide charge: 2+ and 3+
- Monoisotopic: Average:
- Data file: Choose file | FTgreene.mgf
- Data input: Data URL (http or ftp)
- Data format: Mascot generic
- Instrument: ESI-TRAP
- Decoy:
- Precursor: m/z
- Error tolerant:
- Report top: AUTO hits
- Buttons: Start Search, Reset Form

MASCOT : Quantitation

© 2007-2014 Matrix Science



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

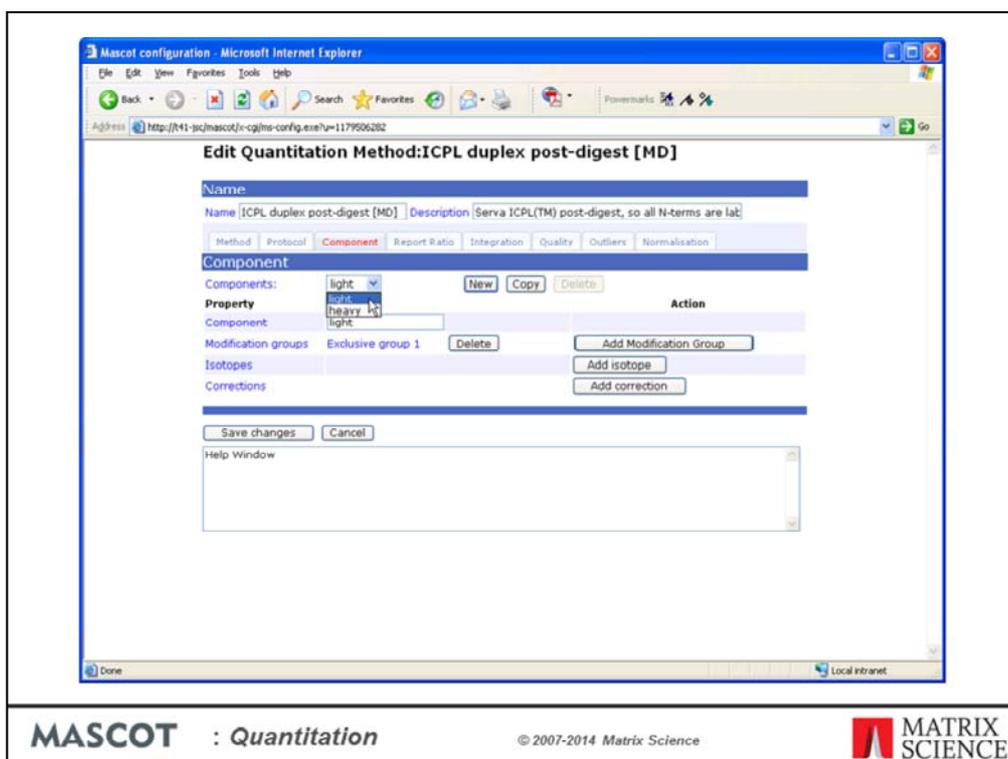
The configuration file that encapsulates the choices and parameters for each quantitation method is called quantitation.xml. This is an XML file, and there is a browser based editor for modifying methods and creating new ones. quantitation.xml lives on the Mascot server and is read by both the search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

Mascot Configuration: Quantitation Methods

Name	Protocol	Copy	Delete	Print
None	null			
iTRAQ 4plex	reporter	Copy	Delete	Print
iTRAQ 8plex	reporter	Copy	Delete	Print
18O corrected multiplex	multiplex	Copy	Delete	Print
SILAC K+6 R+6 multiplex	multiplex	Copy	Delete	Print
TMT 6plex	reporter	Copy	Delete	Print
ICAT ABI Cleavable [MD]	precursor	Copy	Delete	Print
ICPL duplex pre-digest [MD]	precursor	Copy	Delete	Print
ICPL duplex post-digest [MD]	precursor	Copy	Delete	Print
SILAC K+6 R+10 [MD]	precursor	Copy	Delete	Print
18O corrected [MD]	precursor	Copy	Delete	Print
15N Metabolic [MD]	precursor	Copy	Delete	Print

Applied Biosystems iTRAQ(TM) 8-plex reagent

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.

The screenshot shows a web browser window displaying the help page for 'Quantitation: Statistical procedures' on the Matrix Science Mascot database search website. The page includes a navigation menu, a search bar, and a main content area with text and lists of sections and more info.

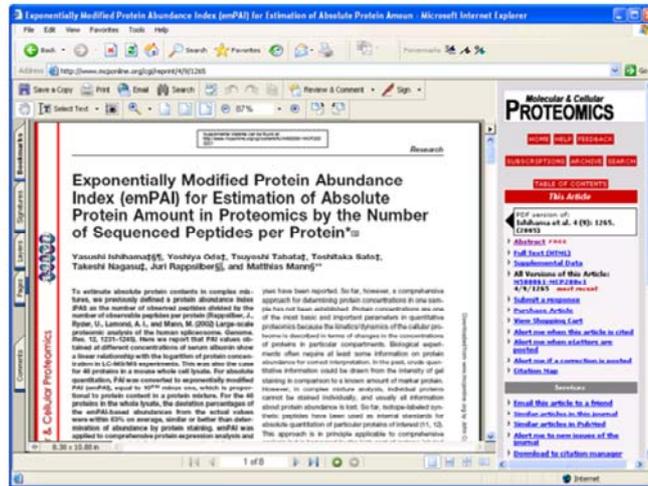
MASCOT : Quantitation

© 2007-2014 Matrix Science

MATRIX SCIENCE

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



MASCOT : Quantitation

© 2007-2014 Matrix Science



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. [PP2B1_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
<input checked="" type="checkbox"/> 27	462.6807	923.3468	923.5116	-0.1649	0	33	0.25	1	R.FPTVALSK.Y
<input checked="" type="checkbox"/> 41	517.1760	1032.3375	1032.5604	-0.2229	0	71	6.4e-05	1	R.GSSIFGLAPGR.A
<input checked="" type="checkbox"/> 62	564.6804	1127.3463	1127.5764	-0.2301	0	9	1.2e+02	1	R.GFFLFVEGGR.I

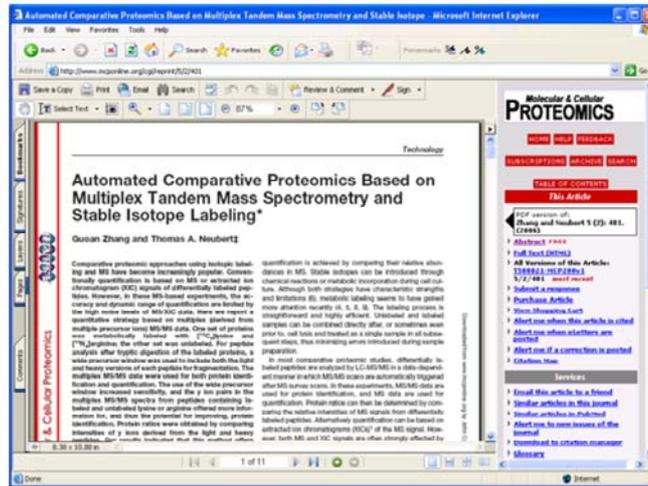
MASCOT : Quantitation

© 2007-2014 Matrix Science



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



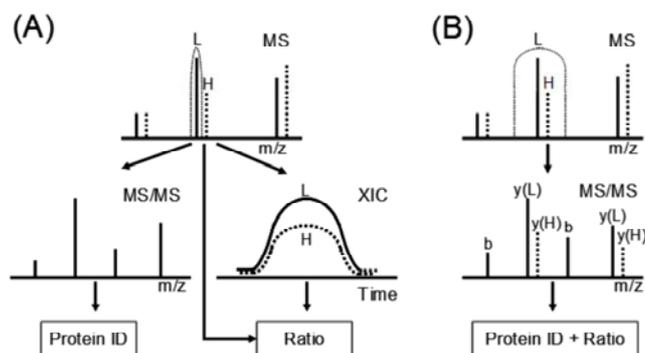
MASCOT : Quantitation

© 2007-2014 Matrix Science



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



MASCOT : Quantitation

© 2007-2014 Matrix Science

MATRIX
SCIENCE

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.

Peptide Summary Report

www.matrixscience.com/cgi/master_results.pl?file=.../data/F981133.dat

Q	Y	Protein	Mass	Score	Matches	Sequences	Ratio	Median	N	SD(geo)	p-value	Peptide
22	55	1144.6219	2287.2292	2287.1485	0.0008	0 (51)	0.057	1	0	1.886	0.937	K.FGQIVNLDLH ₈ + 13C6_M ₈ (C-term K)
22	55	1147.8098	2293.7258	2293.1686	0.5565	0	51	0.853	1	1.321	0.00987	K.VEGSEPTTFNLEFIGNLNPW ₅ + 13C6_M ₈ (C-term K)

7. **EPR2_MOUSE** Mass: 117417 Score: 140 Matches: 4(1) Sequences: 4(1)
Ephrin type-B receptor 2 OS=Mus musculus GI=12982 PE=1 SV=3

Quantitation: Ratio Median N SD(geo) p-value
Heavy/Light 10.965 2 1.054 0.00987

Query	Observed	M(expt)	M(calc)	Delta Miss Score	Expect	Rank	Unique	Heavy/Light	Peptide		
22	571.1036	1140.1927	1139.6282	0.5645	0	63	0.0058	1	0	10.565	K.FGQIVNLDLH ₈ + 13C6_M ₈ (C-term K)
21	620.9648	1239.9130	1239.6343	0.2807	0	26	2.1e+02	1	0	---	K.VNLTGNTIPDK ₈ + 13C6_M ₈ (C-term K)
21	837.9240	1673.8335	1672.9277	0.9058	0	43	0.45	1	0	11.381	K.AMAPLSSGYNLPLLDL ₇ + 13C6_M ₈ (C-term R)
22	1124.9100	2247.8094	2247.1042	0.7012	0	17	1.2e+02	1	0	---	K.TIPDVTNINVDLHAIK ₈ + 13C6_M ₈ (C-term K)

Proteins matching the same set of peptides:
EPR2_MOUSE Mass: 110688 Score: 140 Matches: 4(1) Sequences: 4(1)
Ephrin type-B receptor 2 OS=Mus musculus GI=12982 PE=1 SV=2

8. **HNRPU_MOUSE** Mass: 90528 Score: 125 Matches: 4(2) Sequences: 2(1)
heterogeneous nuclear ribonucleoprotein U OS=Mus musculus GI=106809 PE=1 SV=6

Quantitation: Ratio Median N SD(geo) p-value
Heavy/Light 0.871 2 1.175 0.4389

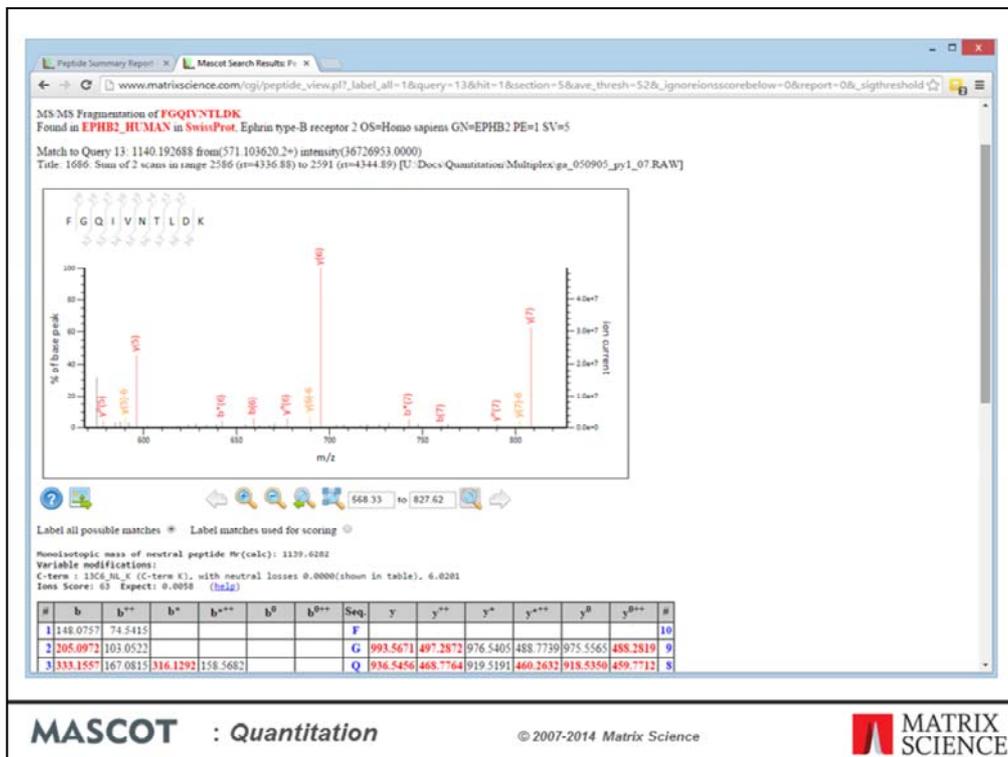
Query	Observed	M(expt)	M(calc)	Delta Miss Score	Expect	Rank	Unique	Heavy/Light	Peptide		
22	692.2000	1382.3854	1381.6911	0.6943	0	27	19	1	0	---	K.VNLTGNTIPDK ₈ + 13C6_M ₈ (C-term K)
28	694.6700	1382.3254	1387.7113	-0.3858	0	(17)	2.1e+02	1	0	---	K.VNLTGNTIPDK ₈ + 13C6_M ₈ (C-term K)
21	824.9751	1647.9357	1646.8376	1.0981	0	(80)	9e-05	1	0	0.976	R.NFILDQTNVSSAAQ ₈ R
22	827.4146	1652.8147	1652.8377	-0.0431	0	104	3.5e-07	1	0	0.777	R.NFILDQTNVSSAAQ ₈ R + 13C6_M ₈ (C-term R)

Proteins matching the same set of peptides:
HNRPU_MOUSE Mass: 87863 Score: 125 Matches: 4(2) Sequences: 2(1)
Heterogeneous nuclear ribonucleoprotein U OS=Mus musculus GI=106809 PE=1 SV=1

9. **SFPQ_MOUSE** Mass: 70102 Score: 122 Matches: 5(2) Sequences: 3(1)
Splicing factor, proline- and glutamine-rich OS=Mus musculus GI=12982 PE=1 SV=2

MASCOT : Quantitation © 2007-2014 Matrix Science **MATRIX SCIENCE**

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.



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

The screenshot shows the 'Edit Quantitation Method: 18O corrected [MD]' configuration page. The 'Name' field is '18O corrected [MD]' and the 'Description' is 'Includes correction for isotope overlap and 95% I'. The 'Component' is set to '1802'. Under 'Corrections', there are three entries:

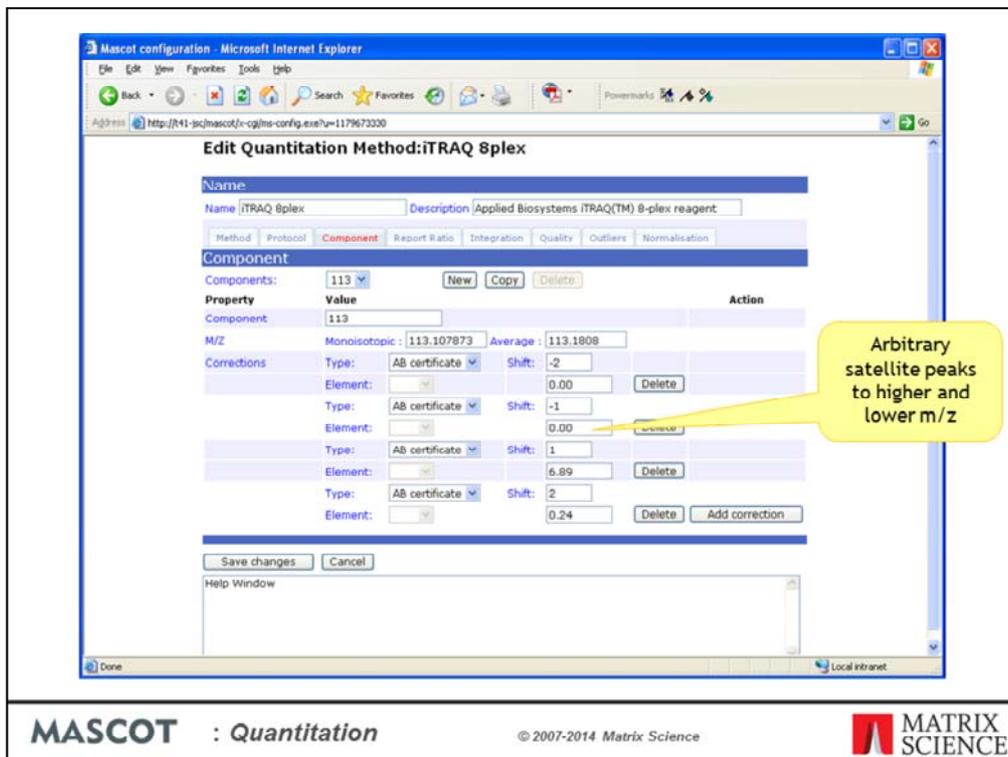
Type	Element	Shift	Action
averagine		0.0	Delete
impurity	O	5.0	Delete
impurity	18O	95.0	Delete, Add correction

Two yellow callout boxes are present:

- One points to the 'averagine' correction with the text: "Satellite peaks to higher m/z due to natural abundance of ¹³C".
- Another points to the 'impurity' correction with the text: "Satellite peaks to lower m/z due to under-enrichment".

At the bottom of the interface, there are 'Save changes' and 'Cancel' buttons, and a 'Help Window'.

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

SPOT_LABEL	SPOT_NAME	SPOT_TYPE	PRECURSOR_MASS	QPMODE
1		Unknown	2336.154	MS-MS 2F
1		Unknown	2617.197	MS-MS 2F
1		Unknown	1252.544	MS-MS 2F
1		Unknown	2514.093	MS-MS 2F
1		Unknown	1344.682	MS-MS 2F
1		Unknown	2463.284	MS-MS 2F
1		Unknown	1230.563	MS-MS 2F
1		Unknown	2479.202	MS-MS 2F
1		Unknown	2447.287	MS-MS 2F
3		Unknown	2542.095	MS-MS 2F
4		Unknown	2496.308	MS-MS 2F

MASCOT : Quantitation

© 2007-2014 Matrix Science



In data processing terms, the reporter protocol is one of the simplest. However, we did find that the peak list exported from the 4000 series data system or submitted to Mascot from GPS Explorer did not have the correct peak areas for the reporter ions. We had to write our own application to export an uncorrected peak list from the Oracle database. We've released this utility, called TS2Mascot, as freeware, and you can download it from our web site.

So, for iTRAQ, we could launch TS2Mascot and choose Mascot Search ...

Quantitation - Reporter

MASCOT MS/MS Ions Search

Your name: ms_user Email: ms@localhost

Search title: NICKMNew Project Spot Set J5C, Job Run 12384, MS-MS 2xV Positive

Database(s): contaminants, CRAP, NCBItr, SwissProt, Uniprot_human Enzyme: Trypsin.P

Allow up to: 2 missed cleavages

Quantitation: iTRAQ 4plex

Taxonomy: Homo sapiens (human)

Fixed modifications: Carbamidomethyl (C) mTRAQ (N-term), mTRAQ (Y), mTRAQ 13C3(15N1) (K), mTRAQ 13C3(15N1) (N-term), mTRAQ 13C3(15N1) (Y), mTRAQ 13C6(15N2) (K), mTRAQ 13C6(15N2) (N-term), mTRAQ 13C6(15N2) (Y), NIPCAM (C), Oxidation (HW), Phospho (ST)

Variable modifications: Oxidation (M)

Peptide tol.: 0.3 Da #13C 0 MS/MS tol.: 0.3 Da

Peptide charge: 1+ Monoisotopic Average

Data input: Data file Choose file CAC12BA.tmp Data URL (http or ftp)

Data format: Mascot generic Precursor: m/z

Instrument: MALDI-TOF-TOF Error tolerant

Decoy Report top: AUTO hits

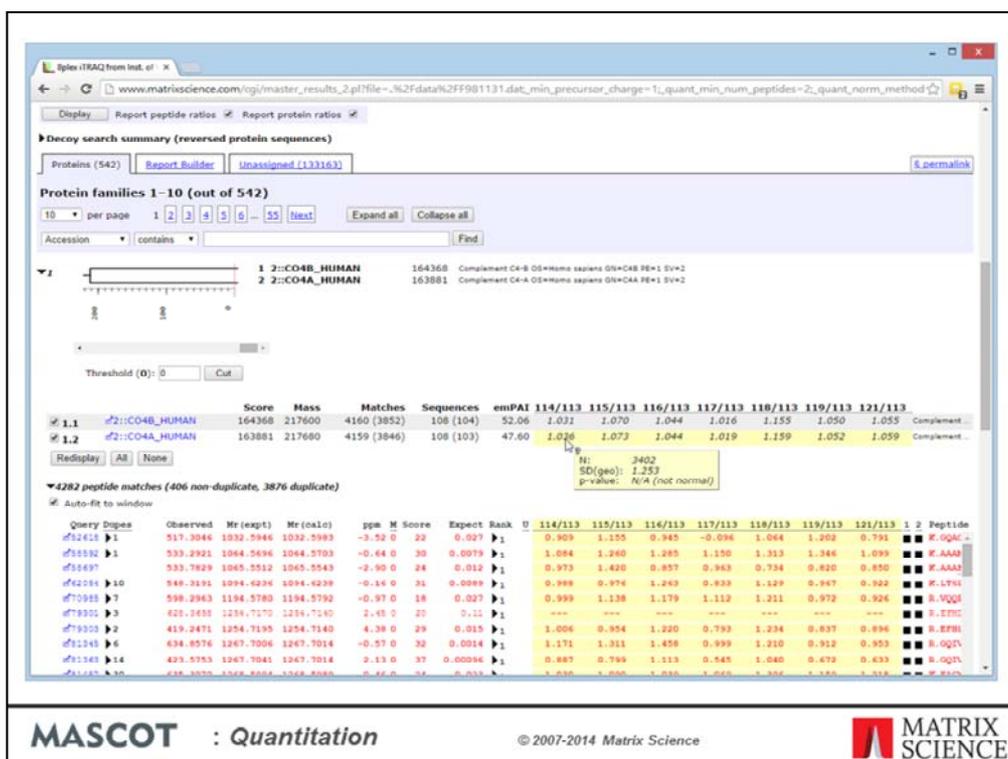
Start Search ... Reset Form

MASCOT : Quantitation

© 2007-2014 Matrix Science



Which brings up 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.

Protein hits (567 proteins)

▼ Columns: Standard (12 out of 58)

Arrangement: <custom> Load Make default

Enabled

- Family
- Member
- Database
- Accession
- Score
- Mass
- Num. of matches
- Num. of significant matches
- Num. of sequences
- Num. of significant sequences
- emPAI
- Description

Available

Protein hits

- Num. of unique sequences
- Num. of significant unique sequences
- Sequence coverage
- pI
- 114/113
- 114/113
- Number of peptides (114/113)
- Significant (114/113) (p-value < 0.05)
- Not-normal (114/113)
- SD(gao) (114/113)
- p-value (114/113)
- 115/113
- 115/113
- Number of peptides (115/113)
- Significant (115/113) (p-value < 0.05)
- Not-normal (115/113)
- SD(gao) (115/113)
- p-value (115/113)
- 116/113

Apply

Filters: (none)

Export as CSV

Family	DB	Accession	Score	Mass	Matches	Match(sig)	Sequences	Seq(sla)	emPAI	Description	
1	1	SwissProt	f2::CO4B_HUMAN	164368	217600	4160	3852	108	104	52.06	Complement C4-B OS=Homo sapiens GN=
1	2	SwissProt	f2::CO4A_HUMAN	163881	217680	4159	3846	108	103	47.60	Complement C4-A OS=Homo sapiens GN=
2	1	SwissProt	f2::APOB_HUMAN	127493	624988	4794	3987	239	218	10.03	Apolipoprotein B-100 OS=Homo sapiens GI=
3	1	SwissProt	f2::CERU_HUMAN	59582	143199	1623	1472	57	50	15.50	Ceruloplasmin OS=Homo sapiens GN=CP P
4	1	SwissProt	f2::A1BG_HUMAN	58871	58330	1557	1532	20	19	11.90	Alpha-1B-glycoprotein OS=Homo sapiens C
5	1	SwissProt	f2::HEMO_HUMAN	44600	58934	2540	1927	33	30	156.06	Hemopexin OS=Homo sapiens GN=HPX PE
6	1	SwissProt	f2::CFAH_HUMAN	37559	167416	1788	1539	67	65	21.74	Complement factor H OS=Homo sapiens GI=

MASCOT : Quantitation © 2007-2014 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.

[Proteins \(542\)](#) | [Report Builder](#) | [Unassigned \(133163\)](#) | [Permalink](#)

Protein hits (555 proteins)

Columns (8 out of 58)

Filters: Database is SwissProt

[Export as CSV](#)

Family	M	Accession	Mass	114/113	N	SD(geo)	Description
365	1	f2::MCPH1_HUMAN	116525	3.595	2	1.930	Microcephalin OS=Homo sapiens GN=MCPH1 PE=1 SV=3
210	1	f2::CRP_HUMAN	29724	2.900	6	1.251	C-reactive protein OS=Homo sapiens GN=CRP PE=1 SV=1
479	1	f2::HAUS7_HUMAN	47480	1.752	2	1.135	HAUS augmin-like complex subunit 7 OS=Homo sapiens GN=HAUS7 PE=1 SV=3
320	1	f2::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	f2::HBB_HUMAN	19731	1.721	33	1.209	Hemoglobin subunit beta OS=Homo sapiens GN=HBB PE=1 SV=2
173	1	f2::HBA_HUMAN	18944	1.694	18	1.234	Hemoglobin subunit alpha OS=Homo sapiens GN=HBA1 PE=1 SV=2
504	1	f2::SNTG2_HUMAN	0	1.634	2	1.983	
323	1	f2::TXNDC5_HUMAN	56364	1.500	2	1.671	Thioredoxin domain-containing protein 5 OS=Homo sapiens GN=TXNDC5 PE=1 SV=2
126	1	f2::HPT_HUMAN	56680	1.494	25	1.245	Haptoglobin OS=Homo sapiens GN=HPT PE=1 SV=1
229	1	f2::PAI1_HUMAN	51161	1.491	2	1.377	Plasminogen activator inhibitor 1 OS=Homo sapiens GN=SERPINE1 PE=1 SV=1
162	2	f2::DEST_HUMAN	25554	1.457	2	2.412	Destin OS=Homo sapiens GN=DEST1 PE=1 SV=3
358	1	f2::CAC2D1_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	f2::GGH_HUMAN	43564	1.408	3	1.332	Gamma-glutamyl hydrolase OS=Homo sapiens GN=GGH PE=1 SV=2
161	1	f2::NGAL_HUMAN	27883	1.383	3	1.027	Neutrophil gelatinase-associated lipocalin OS=Homo sapiens GN=LCN2 PE=1 SV=2
207	1	f2::COTL1_HUMAN	20286	1.381	2	1.047	Coactosin-like protein OS=Homo sapiens GN=COTL1 PE=1 SV=3
302	1	f2::HPSE_HUMAN	73508	1.377	2	1.345	Heparanase OS=Homo sapiens GN=HPSE PE=1 SV=2
286	1	f2::HEM2_HUMAN	40594	1.355	2	2.183	Delta-aminolevulinic acid dehydratase OS=Homo sapiens GN=ALAD PE=1 SV=1
88	1	f2::S100A9_HUMAN	16930	1.351	27	1.183	Protein S100-A9 OS=Homo sapiens GN=S100A9 PE=1 SV=1
291	1	f2::RSLAA_HUMAN	0	1.350	5	1.136	
510	1	f2::VPS53_HUMAN	96078	1.334	2	1.224	Vacuolar protein sorting-associated protein 53 homolog OS=Homo sapiens GN=VPS53 PE=1 SV=1
264	1	f2::CATAL_HUMAN	69333	1.324	3	1.081	Catalase OS=Homo sapiens GN=CAT PE=1 SV=3
247	1	f2::GDI2_HUMAN	62559	1.314	3	1.306	Rab GDP dissociation inhibitor beta OS=Homo sapiens GN=GDI2 PE=1 SV=2
338	1	f2::BLVRB_HUMAN	25239	1.312	2	1.178	Flavin reductase (NADPH) OS=Homo sapiens GN=BLVRB PE=1 SV=3
193	1	f2::CAMP_HUMAN	24340	1.309	9	1.125	Cathelicidin antimicrobial peptide OS=Homo sapiens GN=CAMP PE=1 SV=1
271	1	f2::GSR_HUMAN	67328	1.287	2	1.110	Glutathione reductase, mitochondrial OS=Homo sapiens GN=GSR PE=1 SV=2
395	1	f2::TALDO_HUMAN	47085	1.286	2	1.116	Transaldolase OS=Homo sapiens GN=TALDO1 PE=1 SV=2
184	1	f2::LCAT_HUMAN	53777	1.256	4	1.159	Phosphatidylcholine-sterol acyltransferase OS=Homo sapiens GN=LCAT PE=1 SV=1
128	1	f2::CA1_HUMAN	34678	1.255	13	1.405	Carbonic anhydrase 1 OS=Homo sapiens GN=CA1 PE=1 SV=2

MASCOT : Quantitation | © 2007-2014 Matrix Science | **MATRIX SCIENCE**

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.

Mascot Search Results

User :
 Email :
 Search title : **6plex TMT data from a collaborative demonstration by Proteome Sciences and Thermo Scientific, (PQD on LTQ XL).**
 MS data file :
 Database :
 Quantitation : TMT 6plex [method details](#)
 : Proteome Sciences sixplex Tandem Mass Tag
 Timestamp : 4 Apr 2008 at 12:45:45 GMT
 Enzyme : Trypsin/P
 Fixed modifications : Carbamidomethyl (C), TMT6plex (N-term), TMT6plex (K)
 Mass values : Monoisotopic
 Protein Mass : Unrestricted
 Peptide Mass Tolerance: ± 2.5 Da
 Fragment Mass Tolerance: ± 0.35 Da
 Max Missed Cleavages : 2
 Instrument type : ESI-TRAP
 Number of queries : 6131

Protein hits	127/126	128/126	129/126	130/126	131/126	
1.880	2.576	4.305	1.958	0.679	OVAL_CHICK	Ovalbumin - Gallus gallus (Chicken)
1.877	2.754	3.881	1.908	0.512	ALBU_BOVIN	Serum albumin precursor - Bos taur
2.027	3.339	4.409	2.621	0.793	TRYP_BOVIN	Cationic trypsin precursor - Bos te
2.227	3.030	4.318	2.291	0.619	LYSC_CHICK	Lysosyme C precursor - Gallus galli
2.082	2.965	4.078	2.022	0.735	MYG_EQUMU	Myoglobin - Equus burchelli (Plains
2.213	3.186	4.211	2.276	0.779	CAH2_BOVIN	Carbonic anhydrase 2 - Bos taurus (
1.708	2.693	3.259	2.461	0.626	G3P_F16	Glyceradenhyde-3-phosphate dehydro

Select Summary Report

Format As Select Summary (protein hits) [Help](#)

MASCOT : Quantitation © 2007-2014 Matrix Science **MATRIX SCIENCE**

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

Normalisation

Quantitate	Protein ratio type	Median	Normalise to	Median ratio	[help]
	Min. precursor charge	1	<input type="radio"/> of all peptides		
	Min. # peptides	2	<input checked="" type="radio"/> of peptides assigned to accession(s)	HEMO_HUMAN	
	Unique peptides only	<input type="checkbox"/>	<input type="radio"/> of peptide sequence(s)		
	Outlier removal	Automatic			
	Peptide threshold	Above homology	0.05		

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**

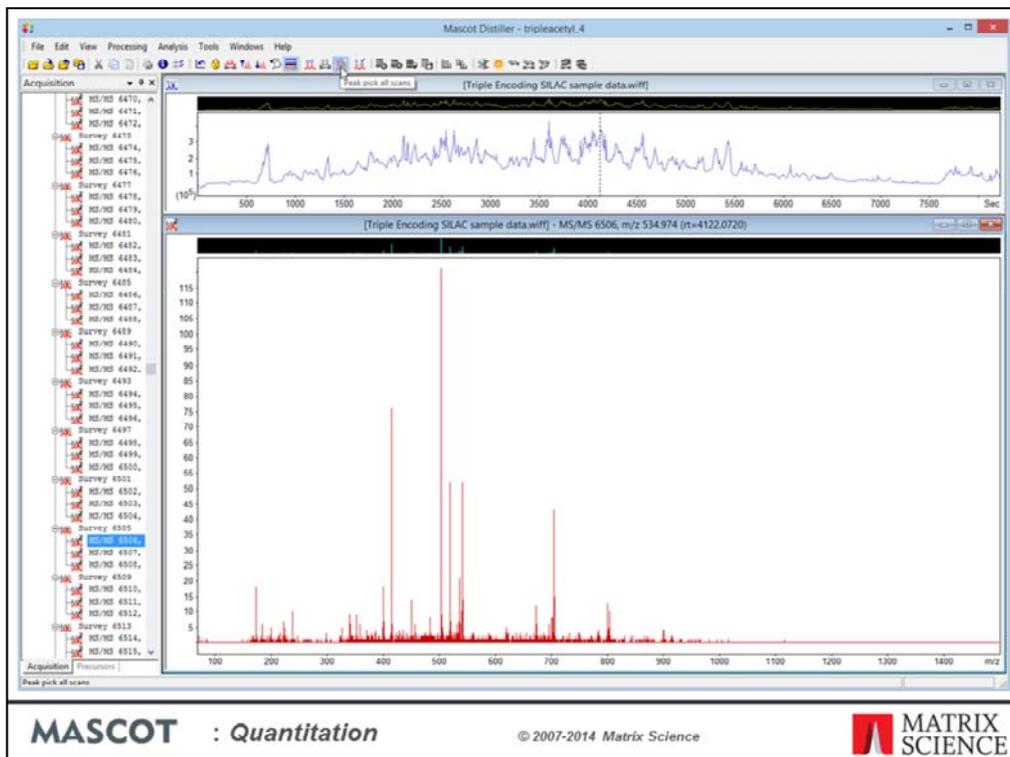
MASCOT : Quantitation

© 2007-2014 Matrix Science

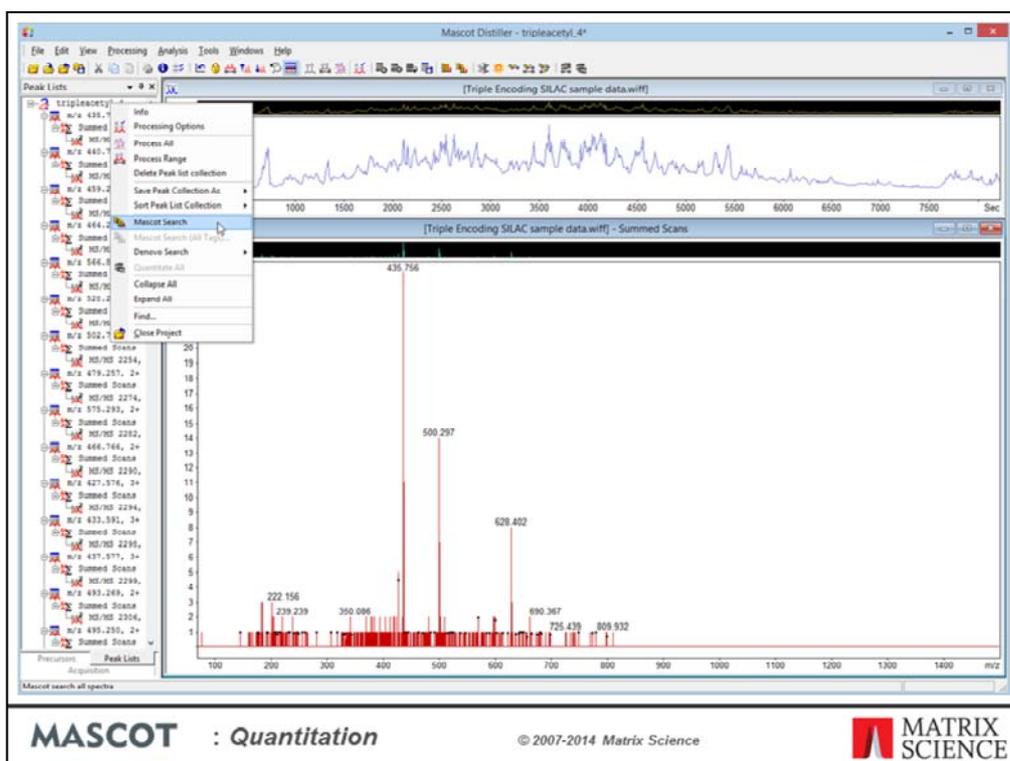


Whether to calculate protein ratios from the average, median or weighted average of the set of peptide ratios is best decided by running some standards (e.g. a cell lysate spiked with varying amounts of a known protein) and seeing which gives the best accuracy and precision.

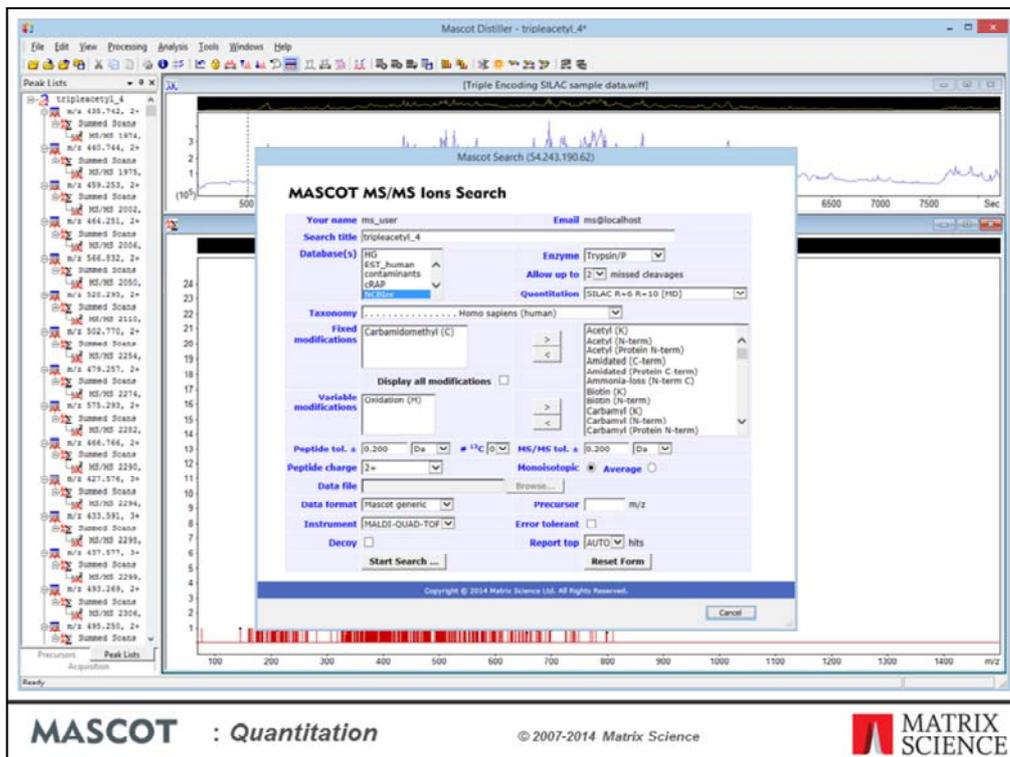
Normalisation is a way to reduce or eliminate systematic errors. In Mascot 2.5, you can normalize to one or more proteins or one or more peptide sequences. Normally, these will have been spiked into the sample for this purpose. You can also perform global normalization by forcing the average or median 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



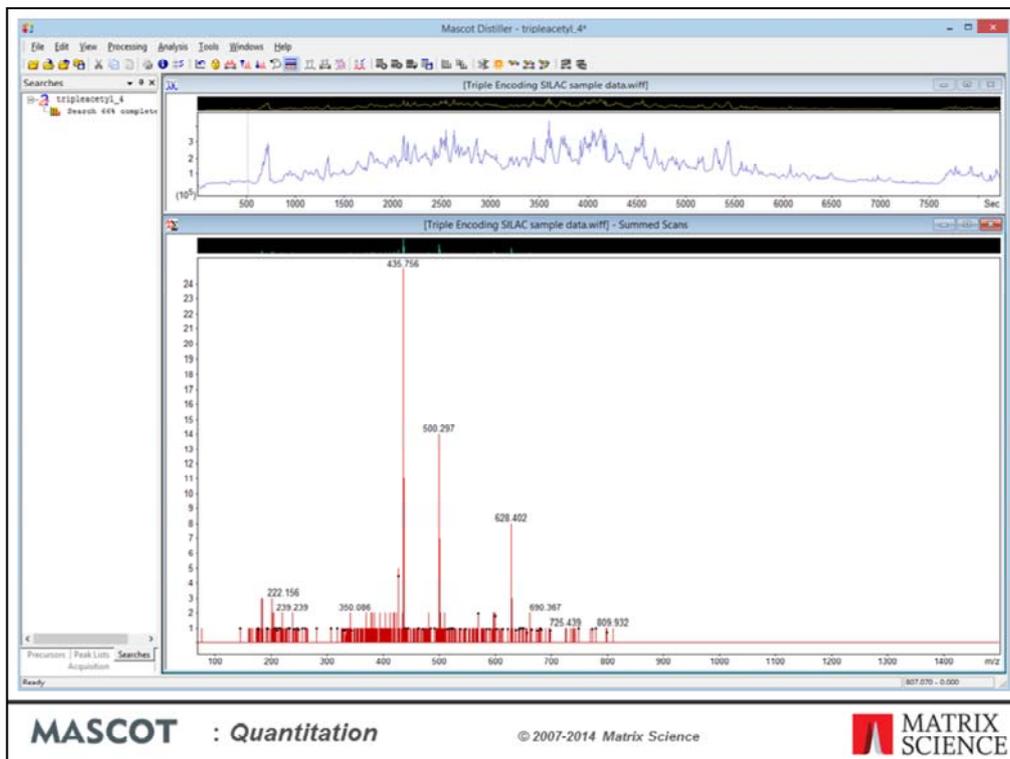
Lets turn to the precursor protocol. This is where we have to use the Mascot Distiller Quantitation Toolbox. I'm going to illustrate the workflow using one of the QStar sample data sets that can be downloaded from the MSQuant home page. This is a SILAC sample with three components: unlabelled, Arg labelled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



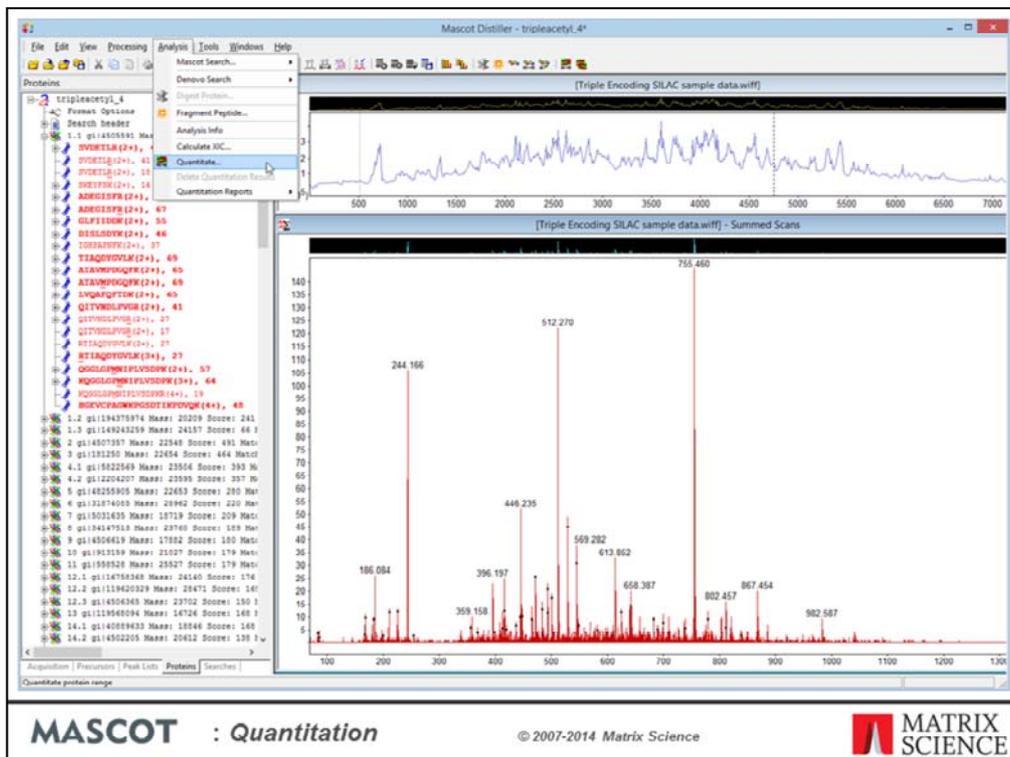
The peak lists are submitted to a Mascot Server to be searched



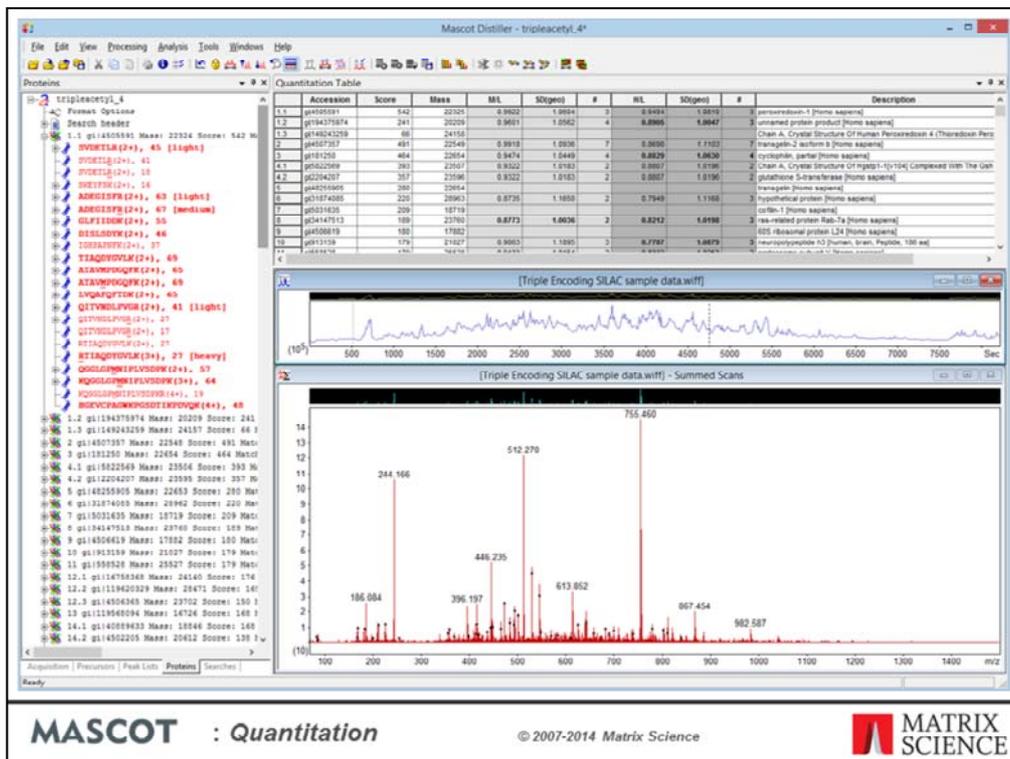
There's an opportunity to tweak the search conditions, but most of the important settings, including the modifications, are embedded in the selected quantitation method



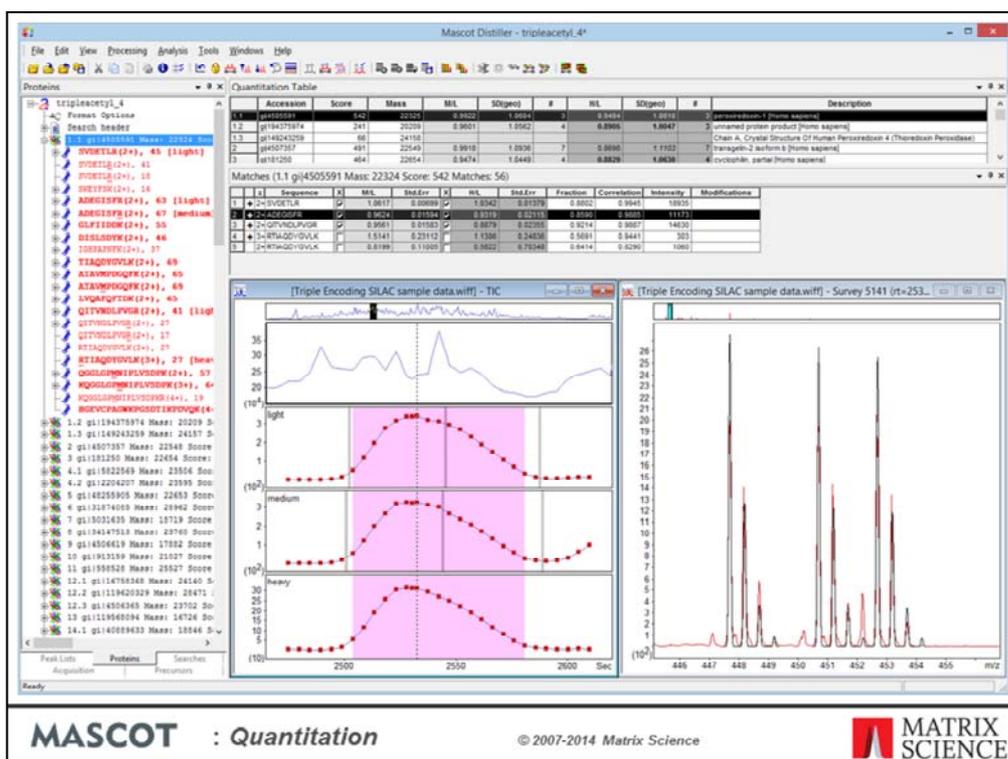
We get progress reports while the search is running



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

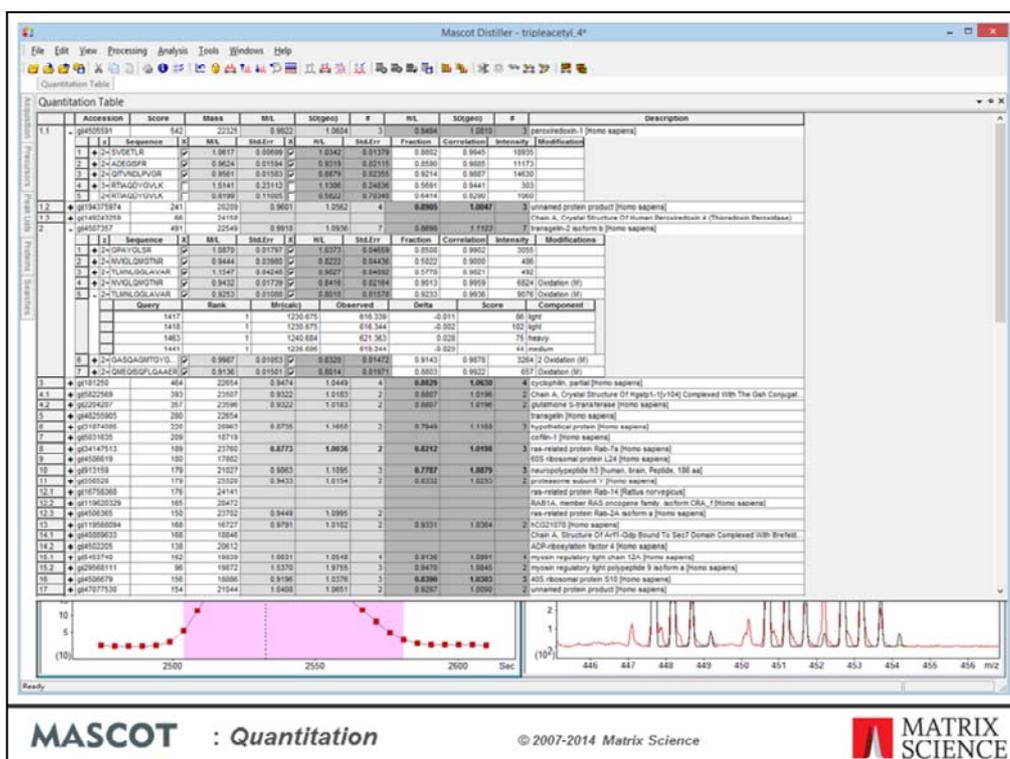


After a few minutes, the quantitation results are displayed as a table.



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.

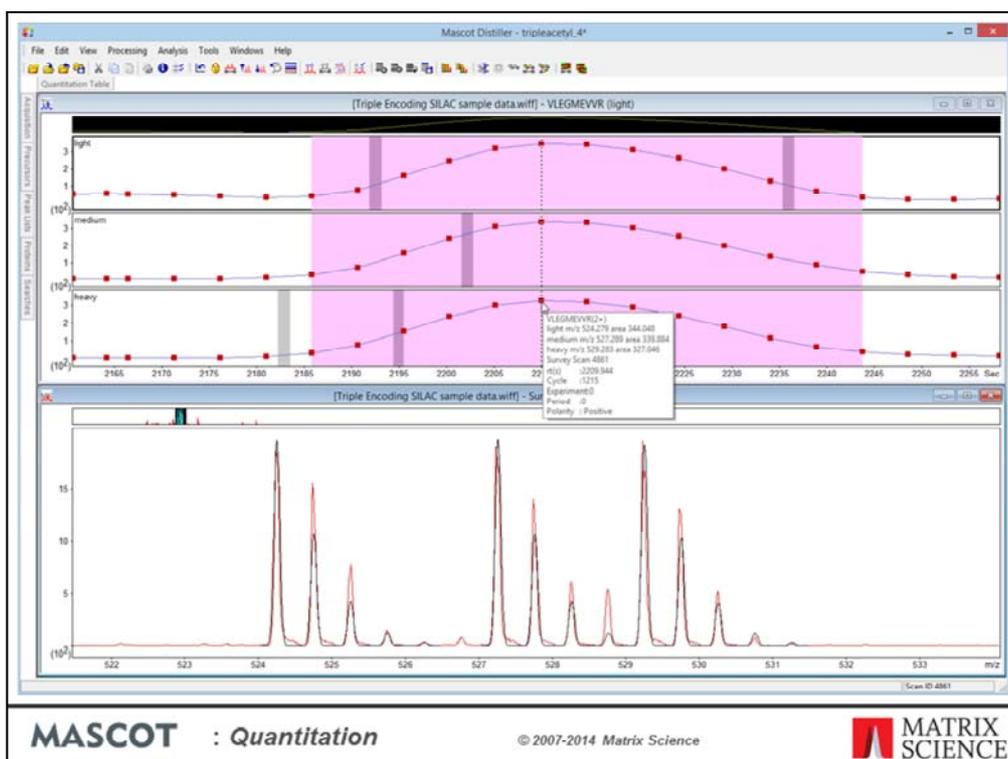
However big your screen, its always difficult to find enough room to display everything. To try and make best use of limited space, the grid and tree can be unpinned, so that they fly out when required and disappear when you move the mouse away.



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.

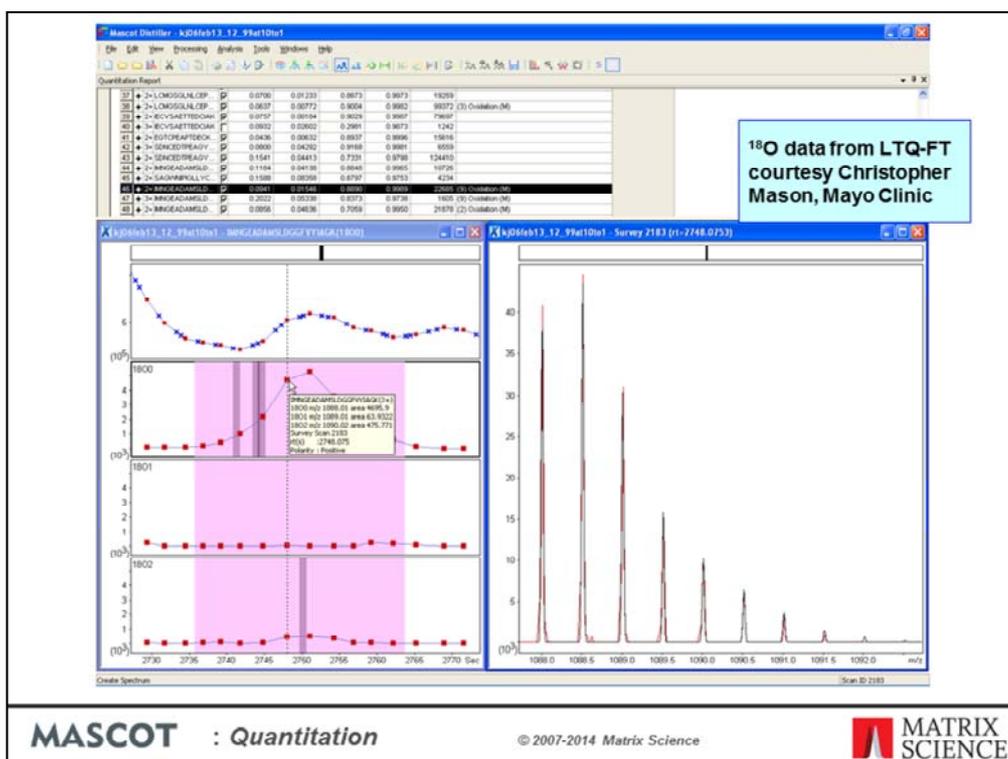
The checkboxes allow you to over-ride the decisions made by the software. If a checkbox is cleared, the peptide ratio is rejected and does not contribute to the protein ratio



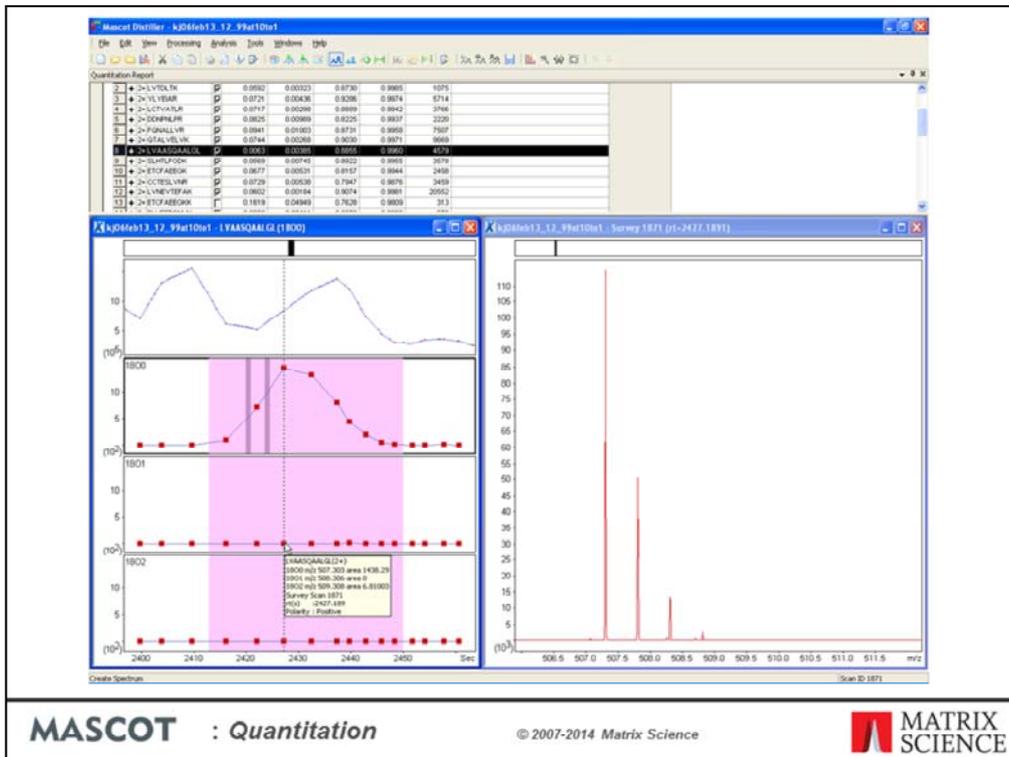
You can spend many happy hours devising different ways to arrange the Windows. With the tree and quant grid unpinned, we can see the chromatogram and scan windows more clearly. A tooltip for each scan provides mass and charge information.

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.

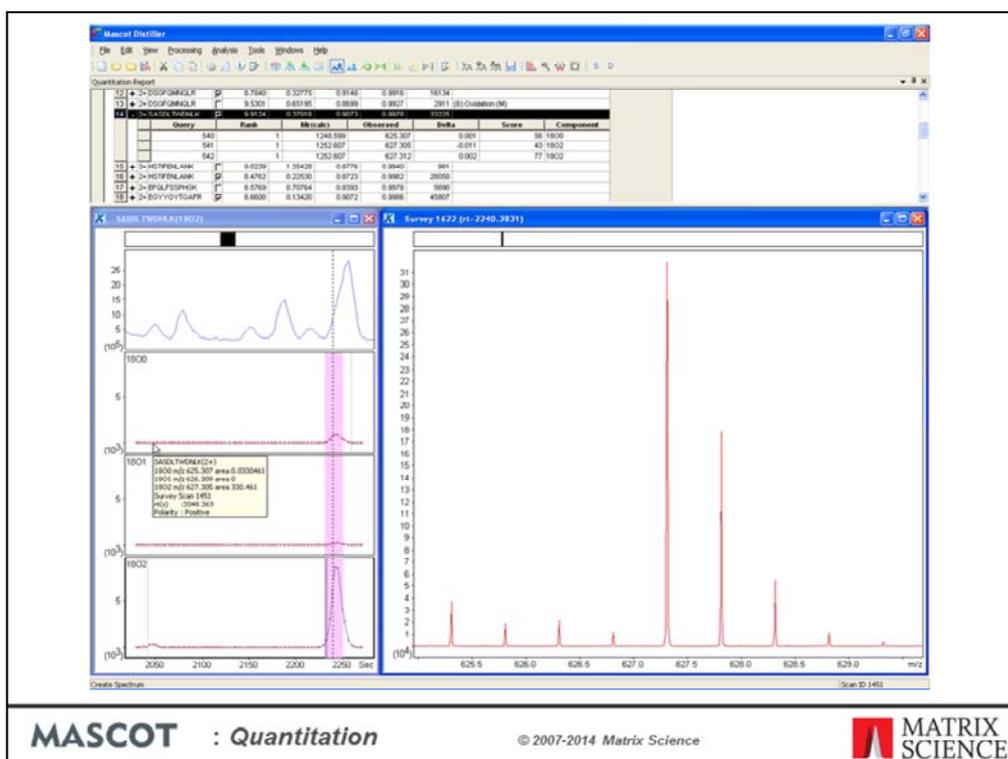
If isotope distributions overlap, this is accounted for in all calculations. In any individual scan, the fit between the experimental and calculated distributions is unlikely to be perfect. Even so, in this case, I hope you'll agree that we can immediately judge that the ratio isn't too bad and we can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



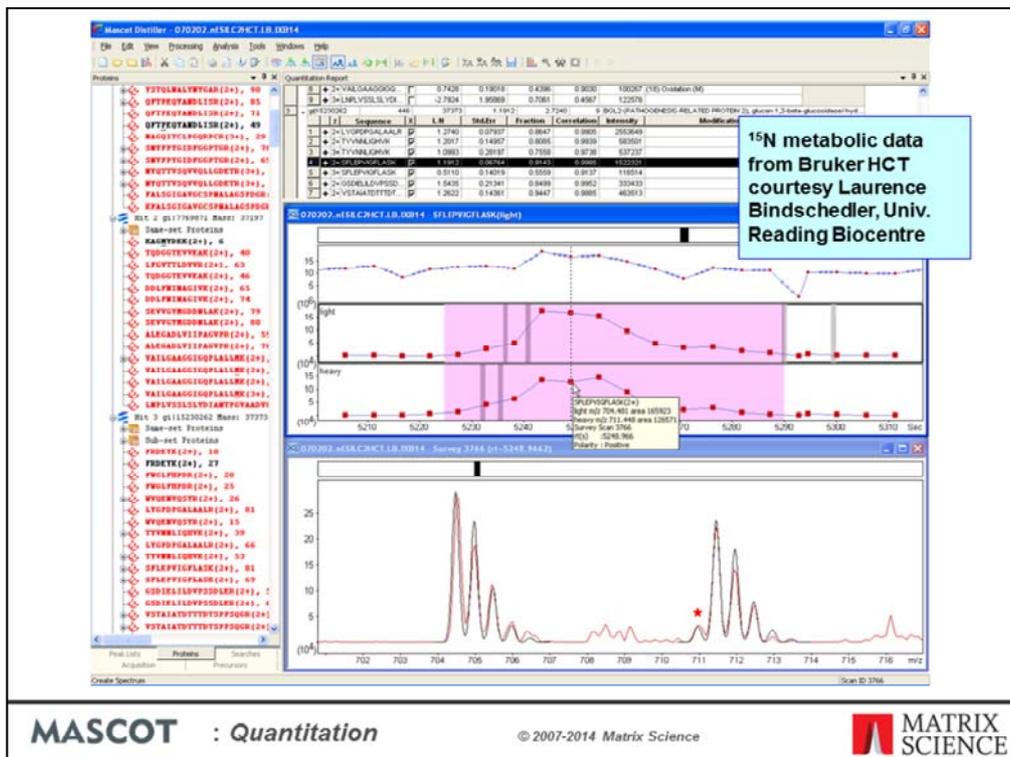
This is O-18 data from an LTQ-FT, courtesy Christopher Mason, Mayo Clinic. The sample is part of a dilution series and this one is particularly tricky because the ratio is 10:1, light to heavy. Obviously, with only a 4 Da separation, the isotope distribution for the heavy is smoothly overlapped by the 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.



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

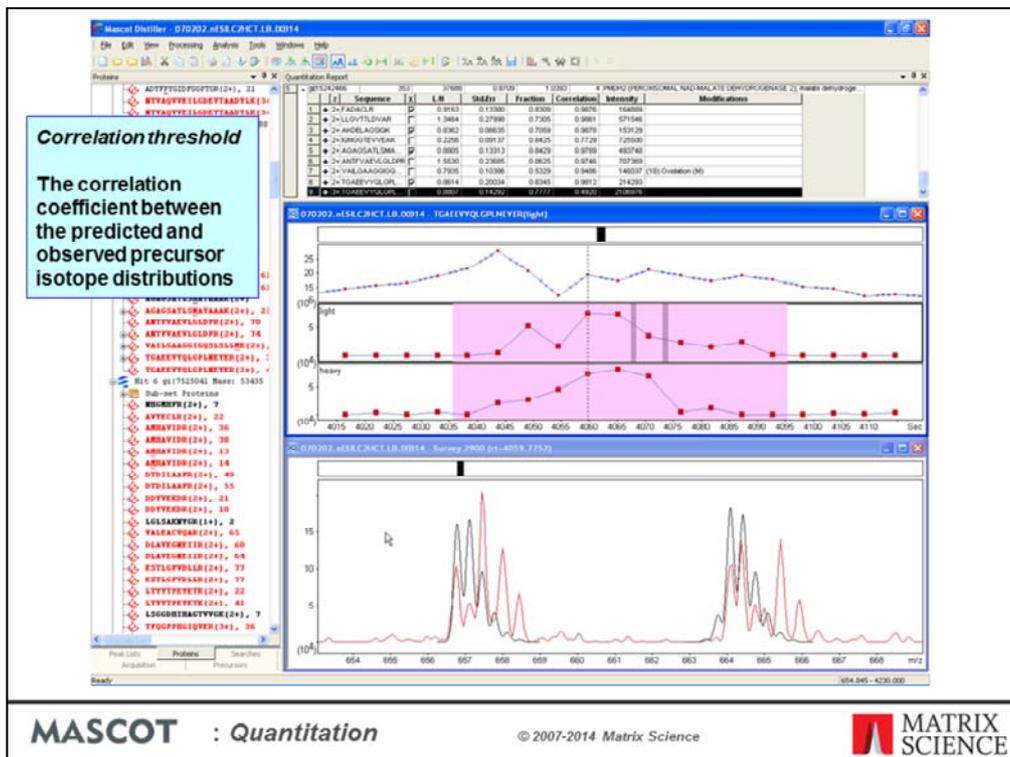


Sometimes, XIC peak detection is a challenge. The starting point is set of scans for which we have Mascot matches. You can hardly see, but in this case, there is a grey band way out here for a match obtained some 3 minutes before the main peak came through. The XIC peak detection algorithm has decided 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.



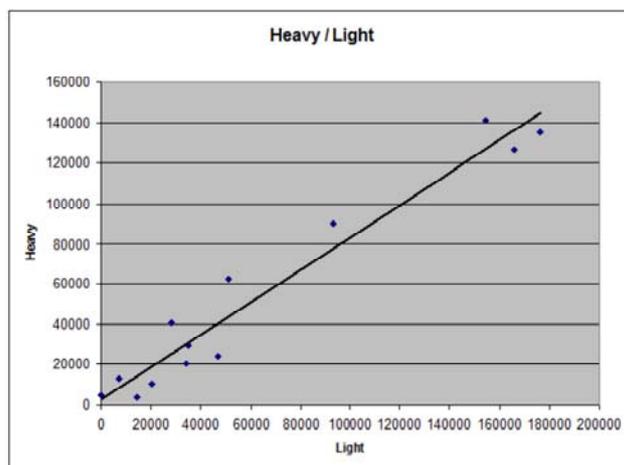
Now some metabolic data, courtesy of Rainer Cramer's group at the Reading Biocentre. In the heavy component, all the nitrogens in the proteins have been replaced by N-15. The calculated distributions (black) fit reasonably well to the observed peaks (red). Notice the peak marked with a red star. This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



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

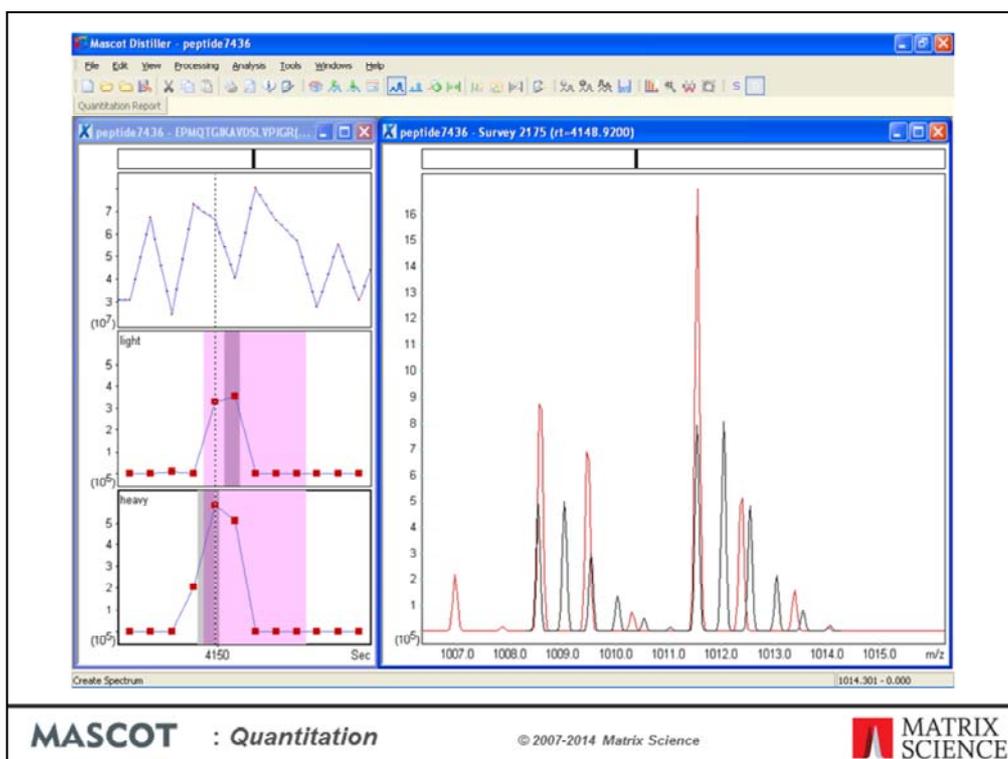
Quantitation - Precursor



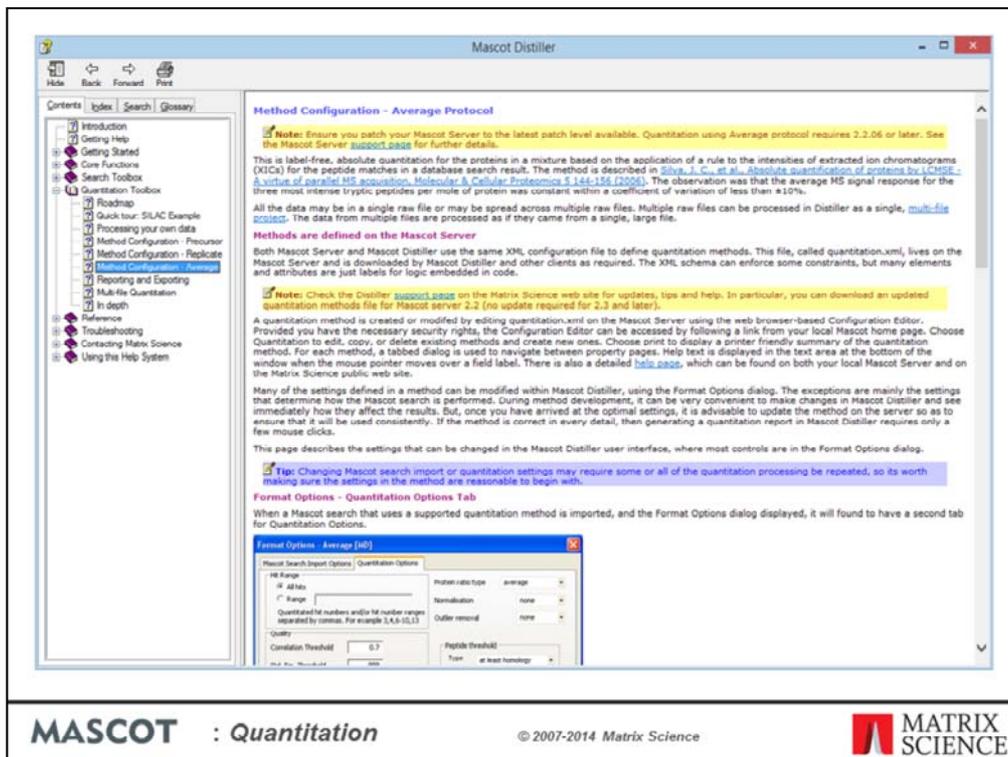
Each point represents the heavy and light intensities in a scan from the XIC peak. The straight line is a least squares fit

- 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.

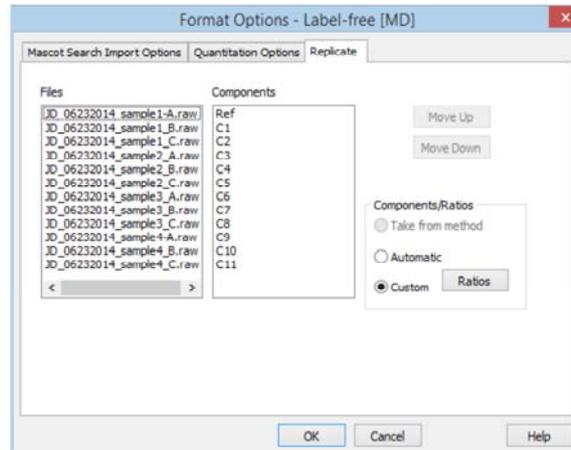


Of course, as always, garbage in means garbage out. Here is a case where the raw data are centroids, not profiles. Although we have good strong MS/MS, when you look at the survey scans, there are no proper isotope distributions. The monoisotopic peaks are approximately in the right place, but the spacings 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



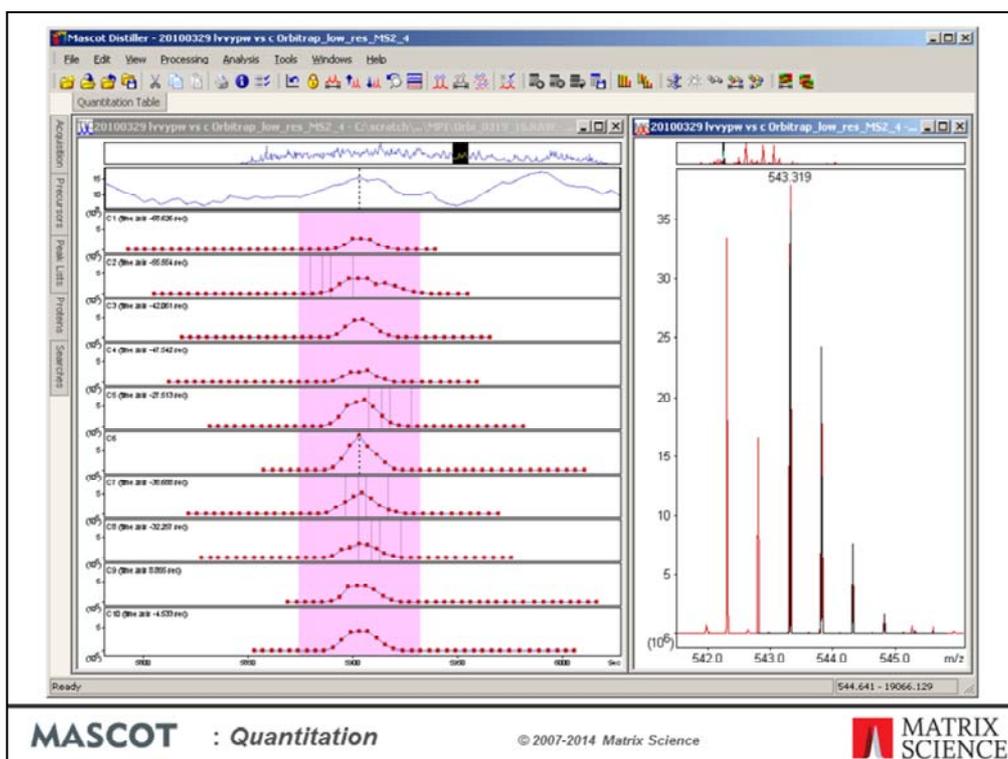
MASCOT : Quantitation

© 2007-2014 Matrix Science

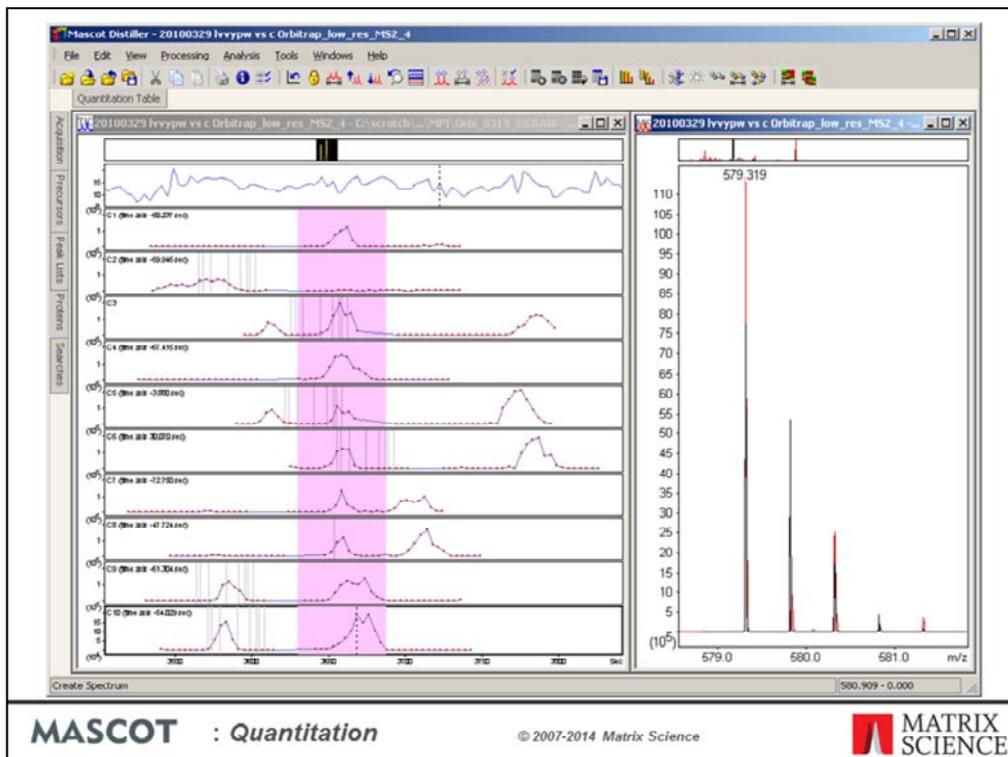


Replicate, you may remember, is 'label free'. Our implementation is identification driven, not feature driven. Distiller starts from the MS/MS data, imports the peptide matches from a Mascot search, then looks in the MS data for the precursors. As with precursor, you don't need MS/MS in every file. You could have one file containing MS/MS and the rest containing just MS and all identifications would be based on the first file.

A replicate project is always a multi-file project, with one file for each component. You can define custom ratios or have them auto-generated, with one file selected as reference and all the others reported relative to it.



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



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.

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
- XSLT style sheets can transform this XML into browser based HTML reports with SVG graphics
- Copying and pasting these reports is best route to 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](#)