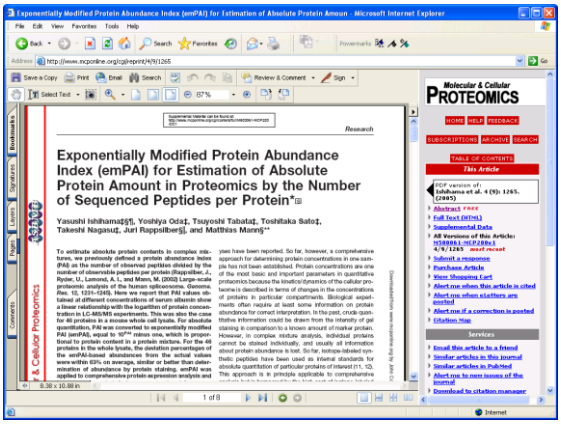
Quantitation – MS2 based methods

In this presentation we will cover quantitation methods that use the data in the peak list, also known as MS2 based methods.

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**

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

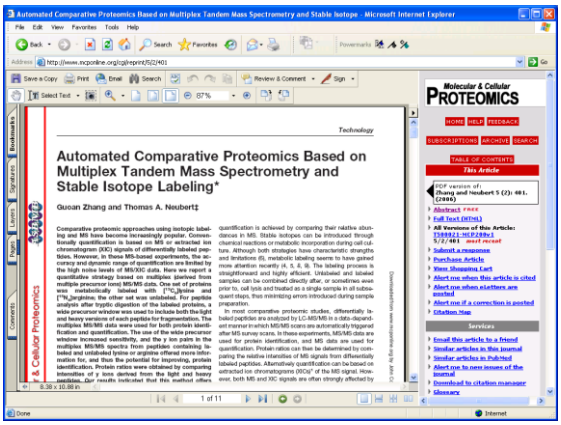
- **Very approximate**
  - Many assumptions in $N_{observed}$ and $N_{observable}$
- **'Always on'**

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

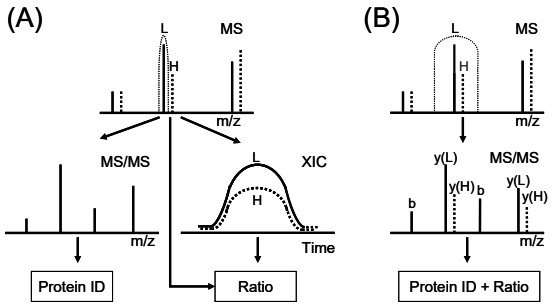
It is very simple. It is also very approximate, because there are many arbitrary assumptions in the way that the number of observed and observable peptides are calculated. Nevertheless, Ishihama's paper shows that it can be a useful guide to relative amounts. emPAI doesn't require a label or special 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

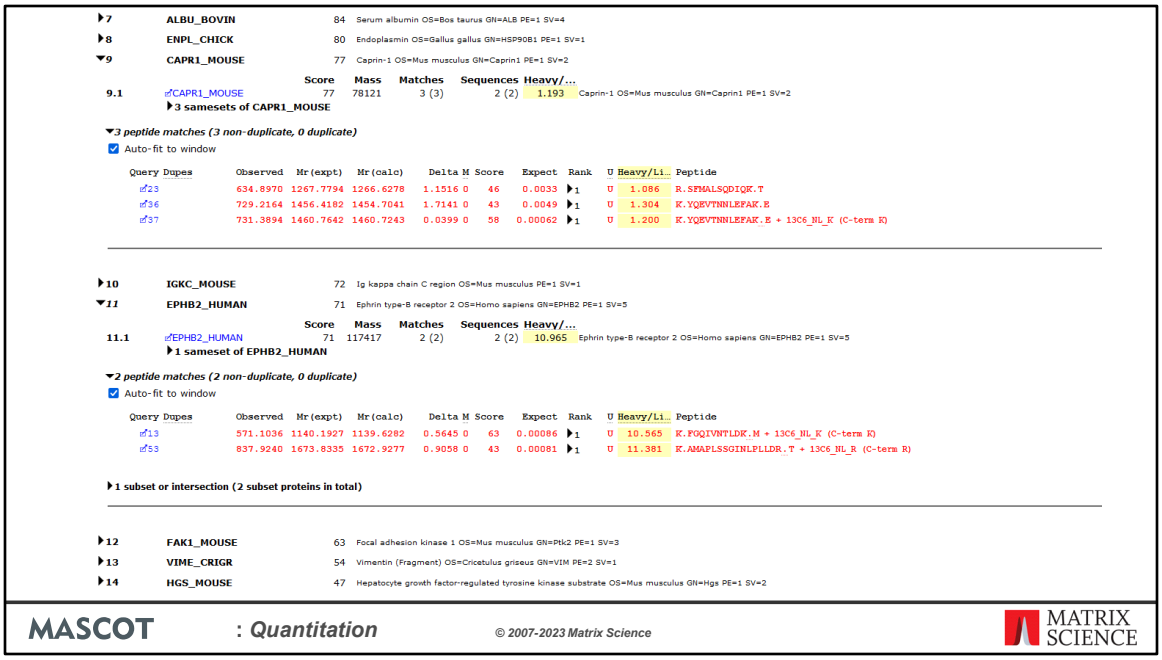
Isobaric Peptide Termini Labeling (IPTL):

- 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.

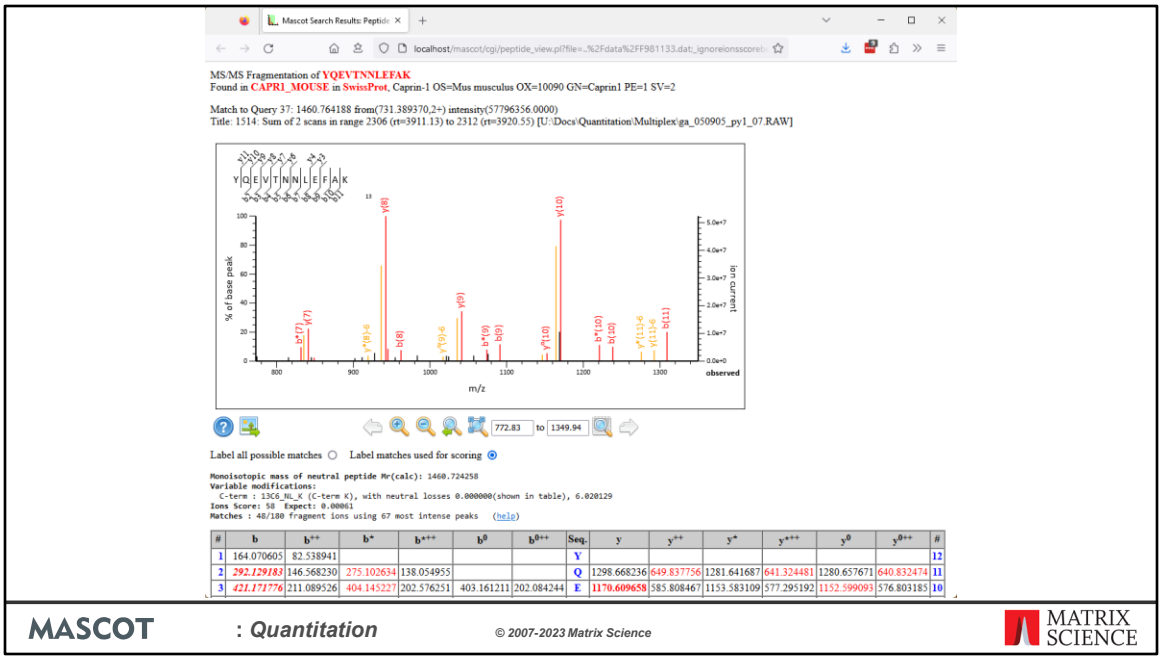
  Component 1:
  Succinyl d0 at the N-term and IMID d4 at C-term

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

Isobaric Peptide Termini Labeling (IPTL) is an 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.

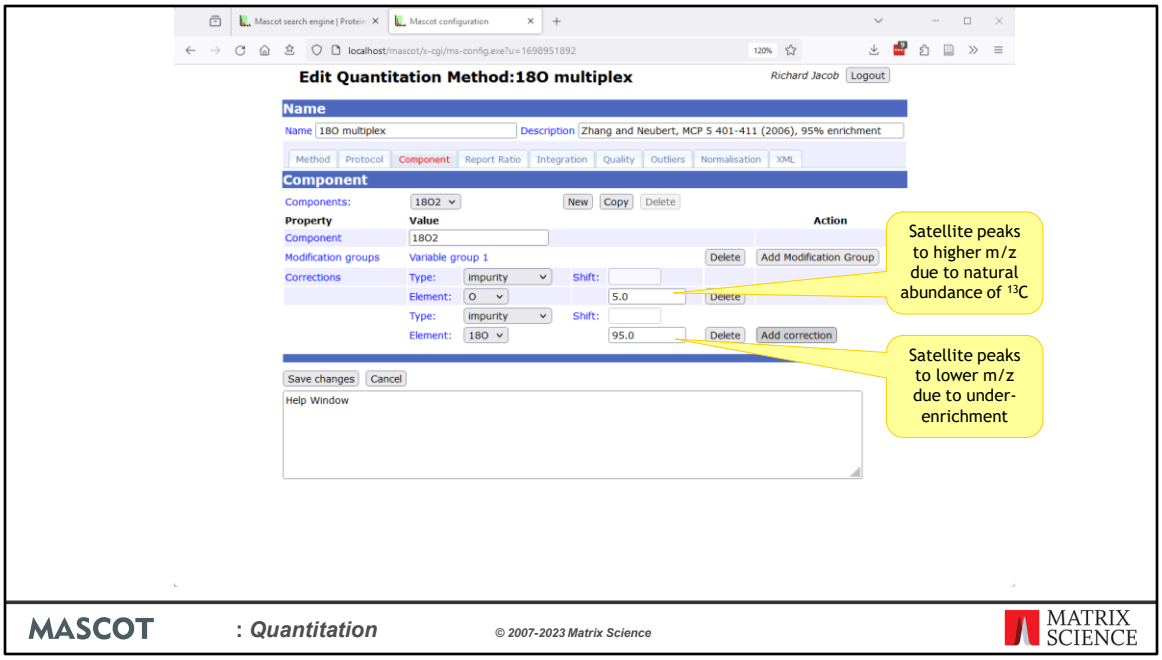


This is an example of Multiplex using a dataset courtesy of Zhang and Neubert. The instrument was an ion trap and the label is 13C(6) SILAC on K and R.

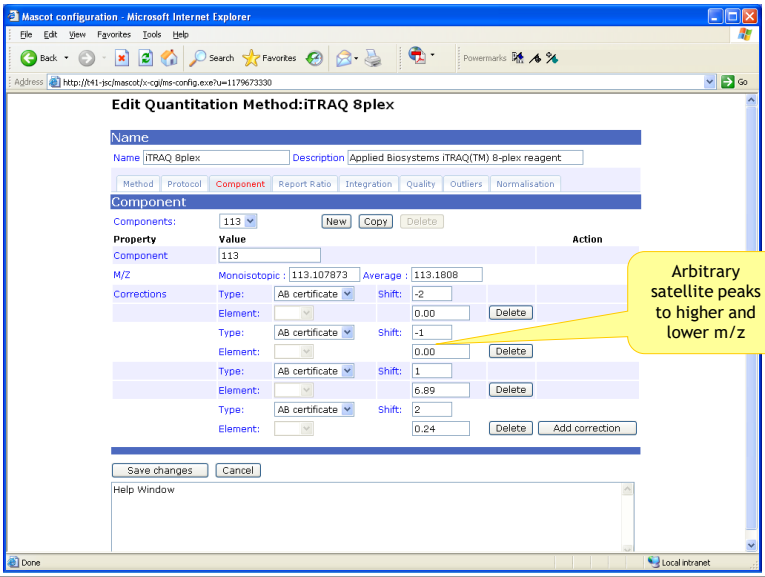


#	b	b^{++}	b^*	b^{*++}	b^0	b^{0++}	Seq.	y	y^{++}	y^*	y^{*++}	y^0	y^{0++}	#
1	164.070605	82.538941					Y							12
2	***292.129183***	146.568230	275.102634	138.054955			Q	1298.668236	649.837756	1281.641687	641.324481	1280.657671	640.832474	11
3	***421.171776***	211.089526	404.145227	202.576251	403.161211	202.084244	E	**1170.609658**	585.808467	1153.583109	577.295192	1152.599093	576.803185	10

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

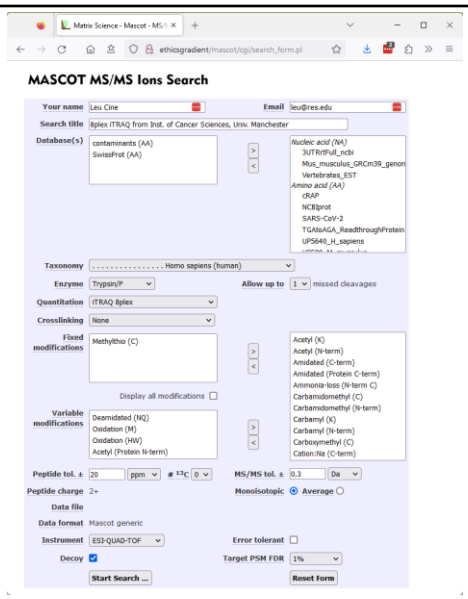


One of the complications of any type of isotope labelling is isotope impurity. It is rarely possible to get 100% enrichment. In the Mascot quantitation schema, this is described by a correction element. An ‘impurity’ correction works "downwards". That is, in this 18O method, some of the intensity of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



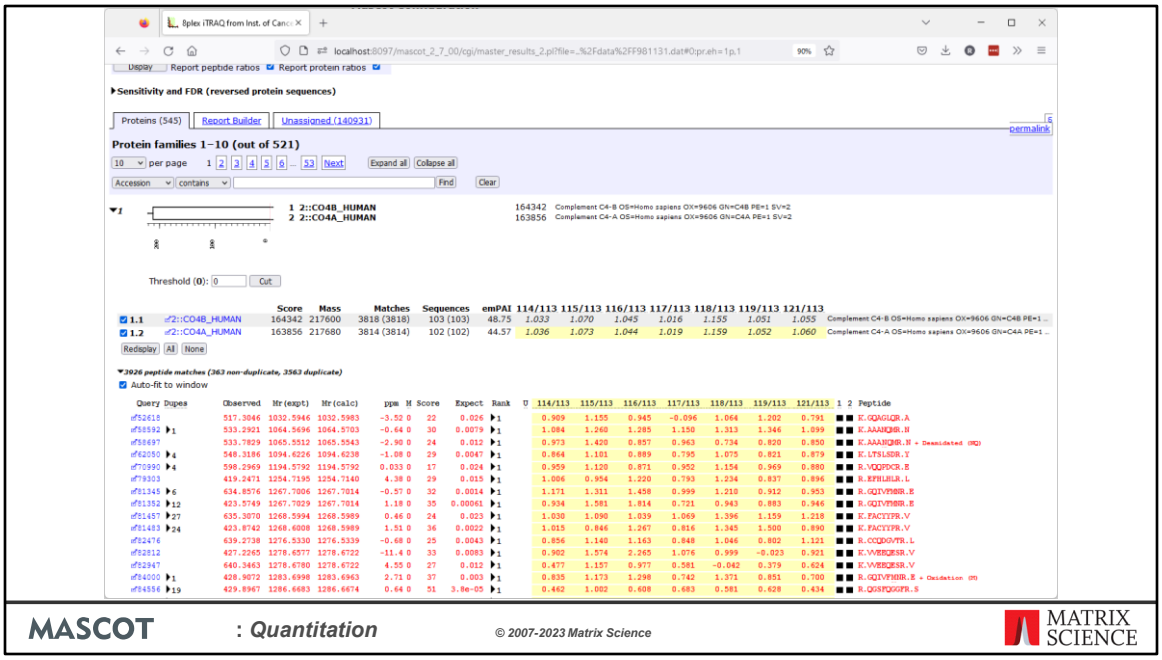
A third type of isotope correction is used in reporter ion technologies like iTRAQ and TMT, where the correction factors are obtained experimentally, by analysing the isolated reagents. This combines both upward and downward corrections for labels which have complex, multi-isotope compositions.

Quantitation - Reporter



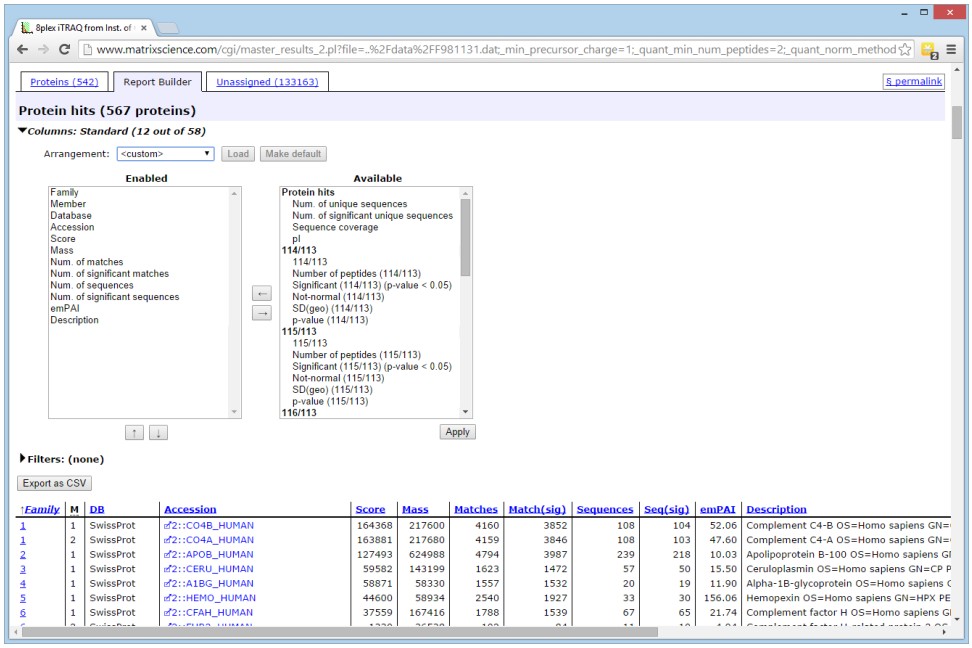
Following on from corrections, let's have a look at reporter ion quantitation. Everything happens on the search engine as the peak list contains all the information required for the quantitation.

Open the search form. We choose an appropriate quantitation method. We don't need to specify the iTRAQ modifications as variable mods because these are pre-defined in the quantitation method. Submit the search…

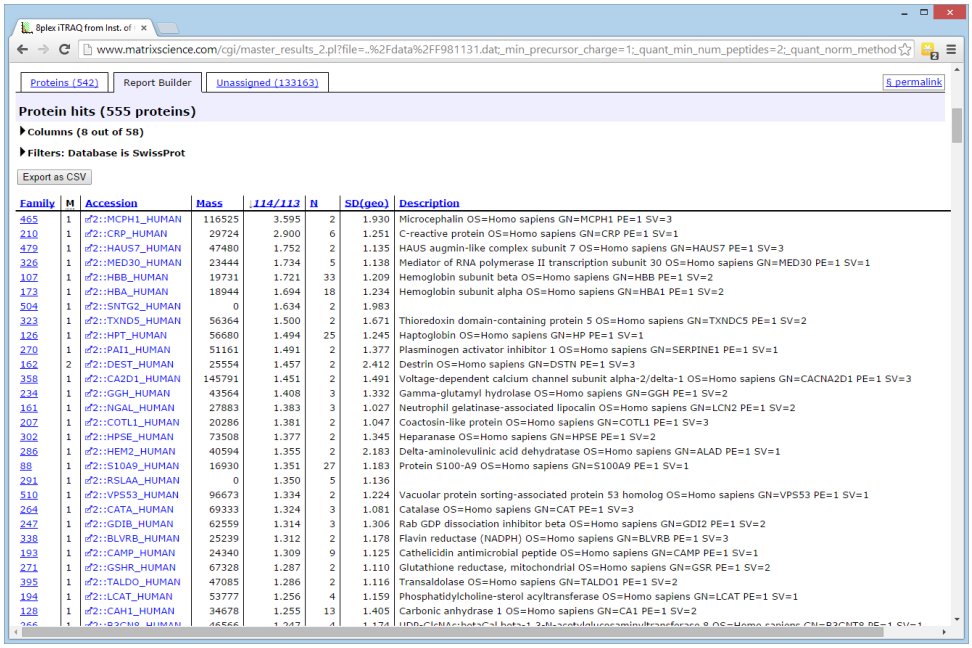


And we have a search report! This example is 8plex iTRAQ. In the Proteins tab, you can toggle display of the ratios for proteins and for individual peptides. For the protein ratios, a tooltip shows the peptide ratio count, the geometric standard deviation, and the p-value for the ratio being different 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.

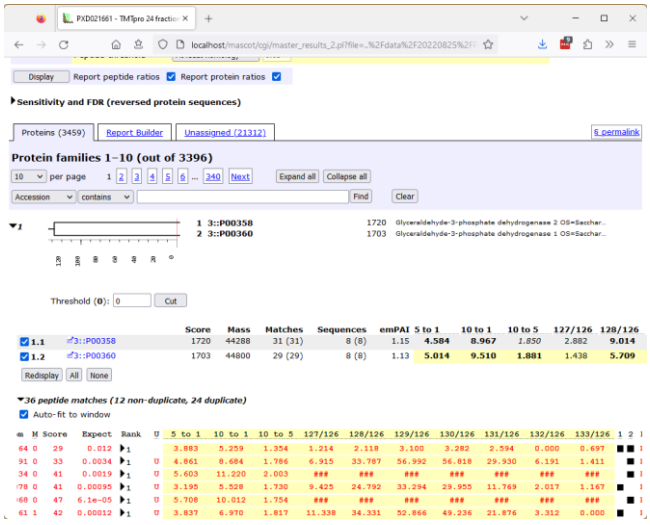


We go to the report builder tab to configure a tabular report covering all the proteins of interest. You can select and re-order the columns, apply filters, and sort the rows.



Family	M	Accession	Mass	↓114/113	N	SD(geo)	Description
465	1	cf2::MCPH1_HUMAN	116525	3.595	2	1.930	Microcephalin OS=Homo sapiens GN=MCPH1 PE=1 SV=3
210	1	cf2::CRP_HUMAN	29724	2.900	6	1.251	C-reactive protein OS=Homo sapiens GN=CRP PE=1 SV=1
479	1	cf2::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	cf2::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	cf2::HBB_HUMAN	19731	1.721	33	1.209	Hemoglobin subunit beta OS=Homo sapiens GN=HBB PE=1 SV=2
173	1	cf2::HBA_HUMAN	18944	1.694	18	1.234	Hemoglobin subunit alpha OS=Homo sapiens GN=HBA1 PE=1 SV=2
504	1	cf2::SNTG2_HUMAN	0	1.634	2	1.983	
323	1	cf2::TXND5_HUMAN	56364	1.500	2	1.671	Thioredoxin domain-containing protein 5 OS=Homo sapiens GN=TXNDC5 PE=1 SV=2
126	1	cf2::HPT_HUMAN	56680	1.494	25	1.245	Haptoglobin OS=Homo sapiens GN=HP PE=1 SV=1
270	1	cf2::PAI1_HUMAN	51161	1.491	2	1.377	Plasminogen activator inhibitor 1 OS=Homo sapiens GN=SERPINE1 PE=1 SV=1
162	2	cf2::DEST_HUMAN	25554	1.457	2	2.412	Destrin OS=Homo sapiens GN=DSTN PE=1 SV=3
358	1	cf2::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	cf2::GGH_HUMAN	43564	1.408	3	1.332	Gamma-glutamyl hydrolase OS=Homo sapiens GN=GGH PE=1 SV=2
161	1	cf2::NGAL_HUMAN	27883	1.383	3	1.027	Neutrophil gelatinase-associated lipocalin OS=Homo sapiens GN=LCN2 PE=1 SV=2
207	1	cf2::COTL1_HUMAN	20286	1.381	2	1.047	Coactosin-like protein OS=Homo sapiens GN=COTL1 PE=1 SV=3
302	1	cf2::HPSE_HUMAN	73508	1.377	2	1.345	Heparanase OS=Homo sapiens GN=HPSE PE=1 SV=2
286	1	cf2::HEM2_HUMAN	40594	1.355	2	2.183	Delta-aminolevulinic acid dehydratase OS=Homo sapiens GN=ALAD PE=1 SV=1
88	1	cf2::S10A9_HUMAN	16930	1.351	27	1.183	Protein S100-A9 OS=Homo sapiens GN=S100A9 PE=1 SV=1
291	1	cf2::RSLAA_HUMAN	0	1.350	5	1.136	
510	1	cf2::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	cf2::CATA_HUMAN	69333	1.324	3	1.081	Catalase OS=Homo sapiens GN=CAT PE=1 SV=3
247	1	cf2::GDIB_HUMAN	62559	1.314	3	1.306	Rab GDP dissociation inhibitor beta OS=Homo sapiens GN=GDI2 PE=1 SV=2
338	1	cf2::BLVRB_HUMAN	25239	1.312	2	1.178	Flavin reductase (NADPH) OS=Homo sapiens GN=BLVRB PE=1 SV=3
193	1	cf2::CAMP_HUMAN	24340	1.309	9	1.125	Cathelicidin antimicrobial peptide OS=Homo sapiens GN=CAMP PE=1 SV=1
271	1	cf2::GSHR_HUMAN	67328	1.287	2	1.110	Glutathione reductase, mitochondrial OS=Homo sapiens GN=GSR PE=1 SV=2
395	1	cf2::TALDO_HUMAN	47085	1.286	2	1.116	Transaldolase OS=Homo sapiens GN=TALDO1 PE=1 SV=2
194	1	cf2::LCAT_HUMAN	53777	1.256	4	1.159	Phosphatidylcholine-sterol acyltransferase OS=Homo sapiens GN=LCAT PE=1 SV=1
128	1	cf2::CAH1_HUMAN	34678	1.255	13	1.405	Carbonic anhydrase 1 OS=Homo sapiens GN=CA1 PE=1 SV=2

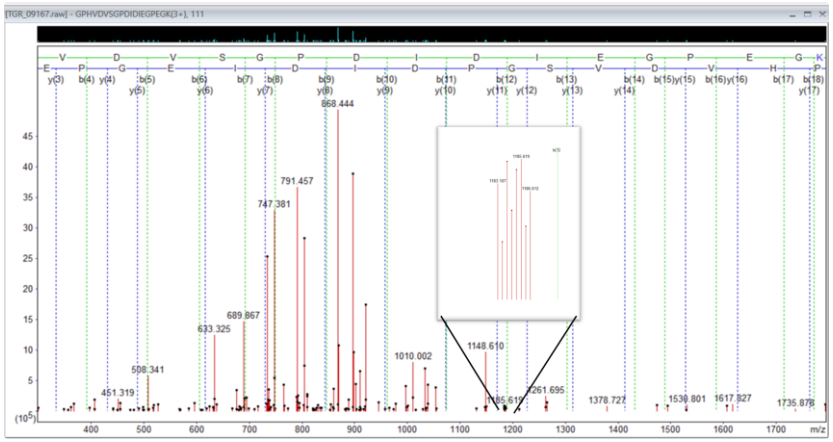
As an example, maybe we want to list proteins with the largest fold change for 114/113 after excluding contaminants. The table has been sorted on descending 114/113.



MASCOT : *Quantitation* MATRIX SCIENCE

The other widely used reporter chemistry is TMT from Thermo Scentific. In this sample the E. coli proteome has been spiked at lower concentration values at ratios of 5 to 1 and 10 to 1 in a background of the human proteome.

TMT and TMTpro complementary ions



MASCOT : *Quantitation* MATRIX SCIENCE

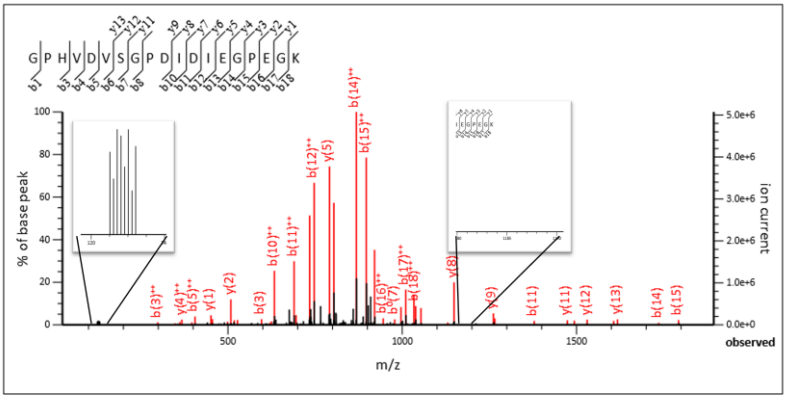
Under certain conditions, TMTpro labeling can generate strong complementary ions. These are the remnants of the labeled peptides after incomplete MS/MS fragmentation, resulting in the loss of the reporter ion and carbon monoxide. The complementary ion consists of the peptide and the balance region of the label.

In this spectrum you can see the expanded doubly charged complementary ion region. If you are observing complementary ions for some spectra, you can remove them with a supplementary script which results in a slight improvement in the peptide scores.

Alternatively, you can tune the fragmentation conditions to promote complementary ions and use them for quantitation. The main reason to use complementary ions for quantitation is that there can be interference in the standard reporter ion region from co-isolating peptides that then fragment along with the target peptide and distort the reporter ion measurements. The complementary ions relate directly to the precursor mass of the peptide so we lessen the interference from other peptides.

The script is run by Mascot Daemon after peak picking but before searching to either remove or move the complementary ions and replace the original reporter ions.

TMT and TMTpro complementary ions



MASCOT : *Quantitation* MATRIX SCIENCE

In this example we are replacing the reporter ions with the complementary ions. The complementary ions are removed from the spectra along with the original reporter ions. New reporter ions are calculated from the complementary ions and inserted back into the peak list. Mascot Server then uses the new reporter ions for quantitation.

There are full instructions on using or removing the complementary ions plus a link to the script in the Mascot Server help pages, quantitation section, reporter ions.

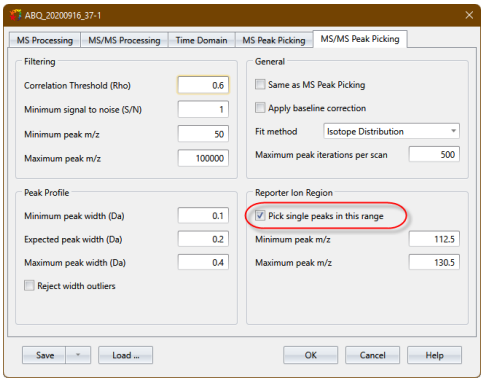
Peak picking

- 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

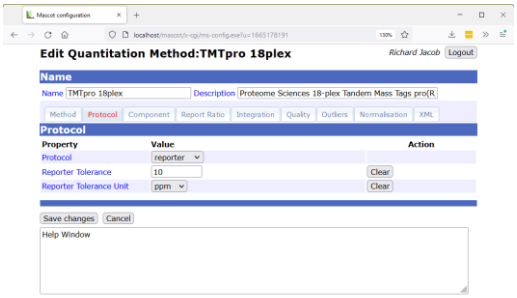
For the reporter protocol, i.e. iTRAQ or TMT, you have to be very careful with peak detection. Reporter ions do not have natural isotope distributions, so anything that assumes this will not be reliable.

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

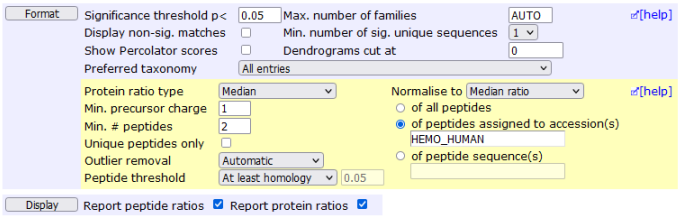


MASCOT : Quantitation MATRIX SCIENCE

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 13C and 15N. 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 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.

Quantitation talks

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

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

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