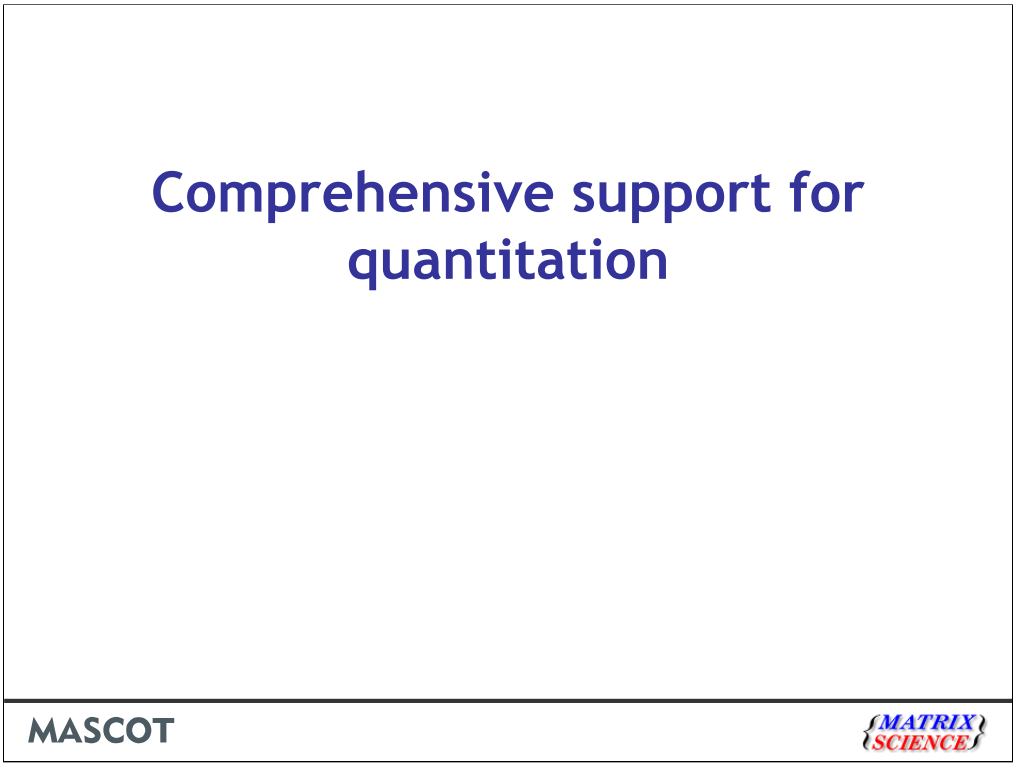


One of the major new features in the current release of Mascot is support for quantitation. This is still work in progress. Our goal is to support all of the popular methodologies.

Quantitation - Overview

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

MASCOT : *Quantitation* 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. If anyone can see a method that doesn't fit to one of these, we'd be very grateful for details.

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.

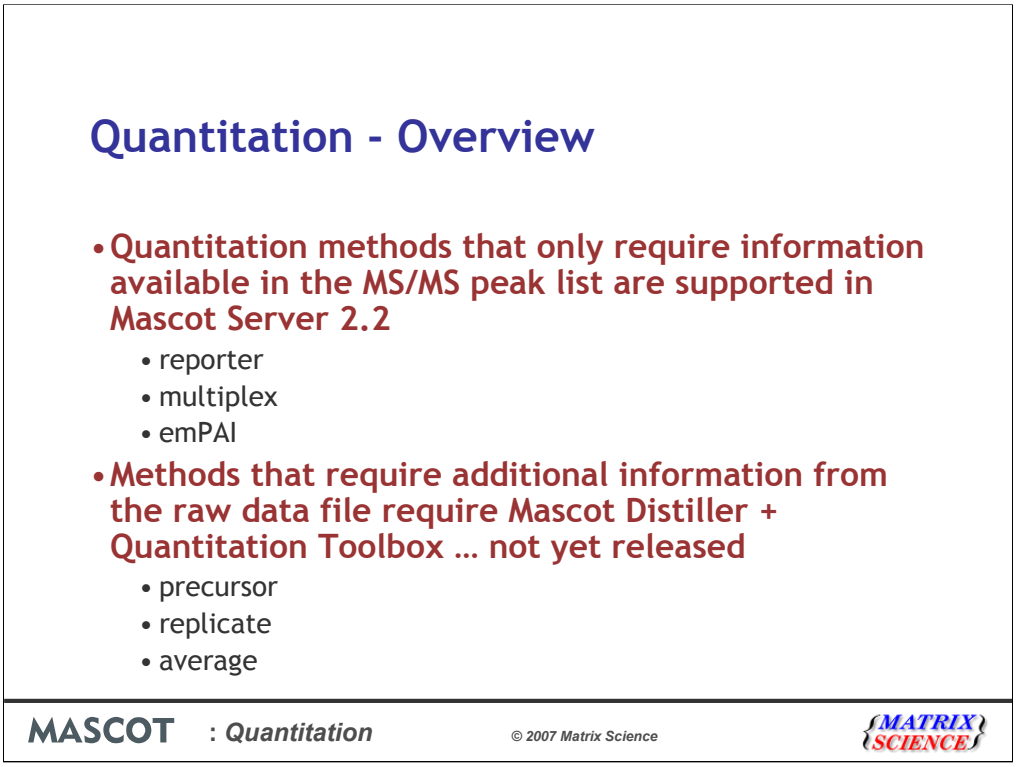
Multiplex is quantitation based on the relative intensities of sequence ion fragment peaks within an MS/MS spectrum. This is a novel approach, which can be used with any chemistry that labels one peptide terminus, creating a small mass shift, such as 18O or SILAC under certain conditions.

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.

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

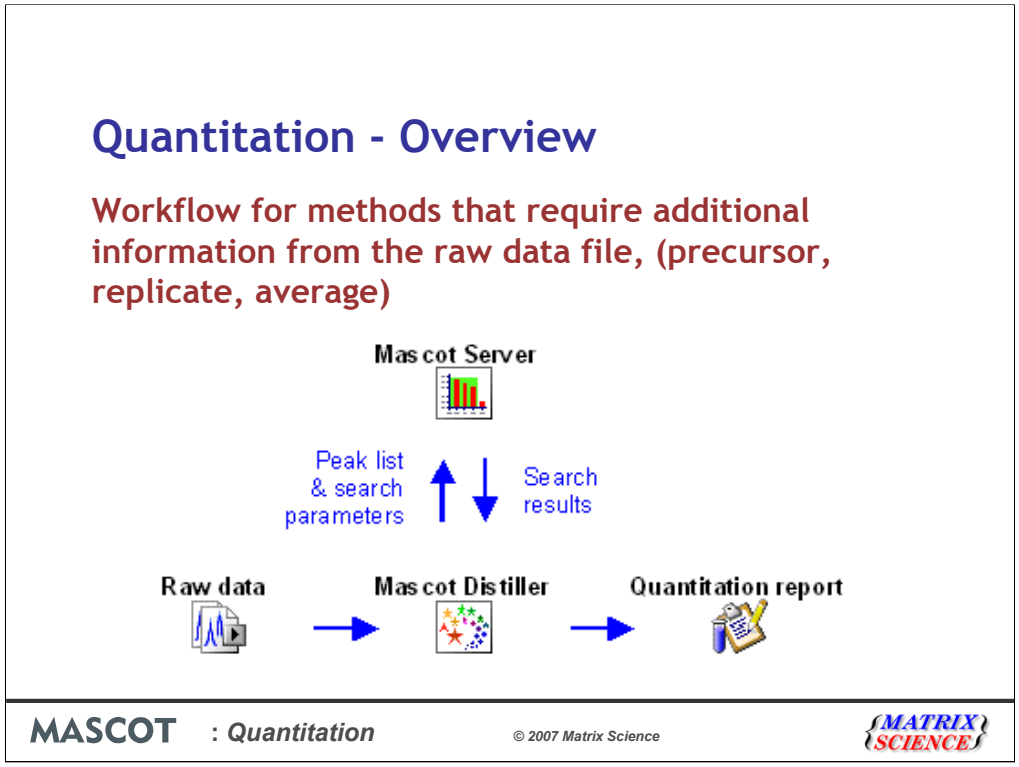
Average is quantitation for the proteins in a mixture based on the application of a rule to the intensities of extracted ion chromatograms (XICs) for the peptide matches in a database search result. For example, the average intensity for the three strongest peptide matches per protein

The rows with a blue background are the protocols that are fully implemented in Mascot 2.2



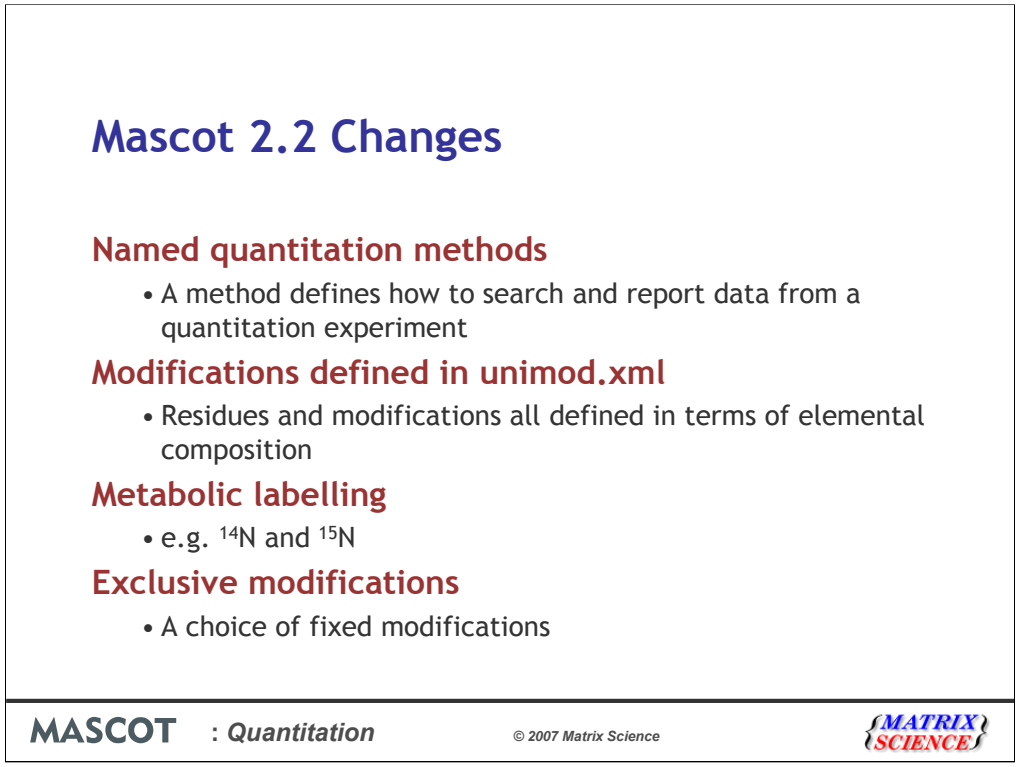
The common factor for these protocols is that all of the information required for quantitation is contained in the peak list.

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.



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.

The Quantitation Toolbox for Distiller is still under development. We will try to get it released as soon as we can.



Let me emphasise that the changes to the Mascot search engine are complete and released.

The set of quantitation methods is defined in a new XML configuration file, called quantitation.xml. As with other configuration files, this file lives on the Mascot Server and is downloaded by Mascot Distiller and other clients as required.

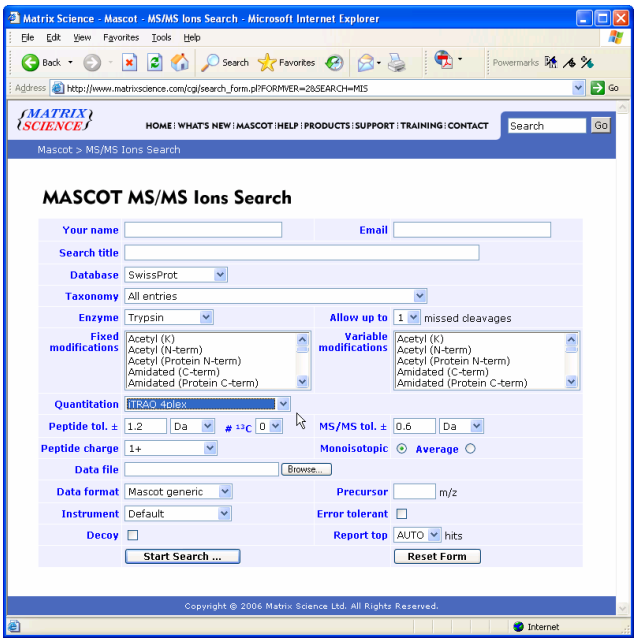
The introduction of quantitation has required changes in the way that modifications are handled. Mascot now takes its modification definitions direct from an XML representation of the Unimod database.

One factor that forced this change was the need to support metabolic labelling, in which the isotopic label is present throughout the peptide backbone. This requires residues and modifications to be defined and manipulated as elemental compositions.

We have also introduced exclusive modifications, which can be thought of as a choice of fixed modifications. In many quantitation experiments, separate samples are derivatised then pooled. Thus, a given peptide may carry one or the other set of modifications, but never a mixture of both. Some people use the term "binary" for this type of specificity. We prefer exclusive because binary implies only two possibilities. The real importance of this is that it keeps the search space small, and avoids the ‘combinatorial explosion’ that can happen with too many variable modifications

Quantitation

Simplicity for user



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 new control, which replaces the old ICAT checkbox.

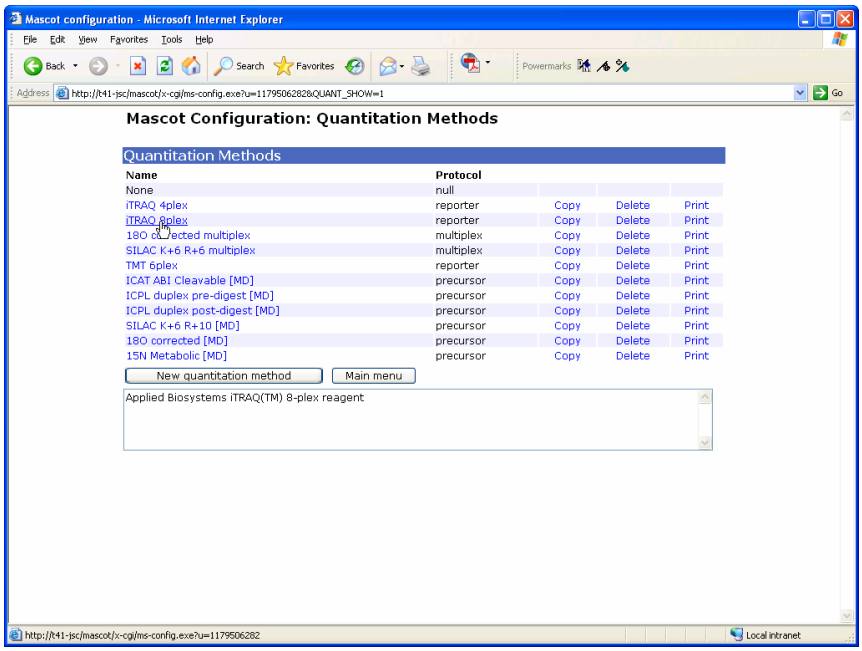
Methods that have [MD] at the end are the ones that require Mascot Distiller

Quantitation

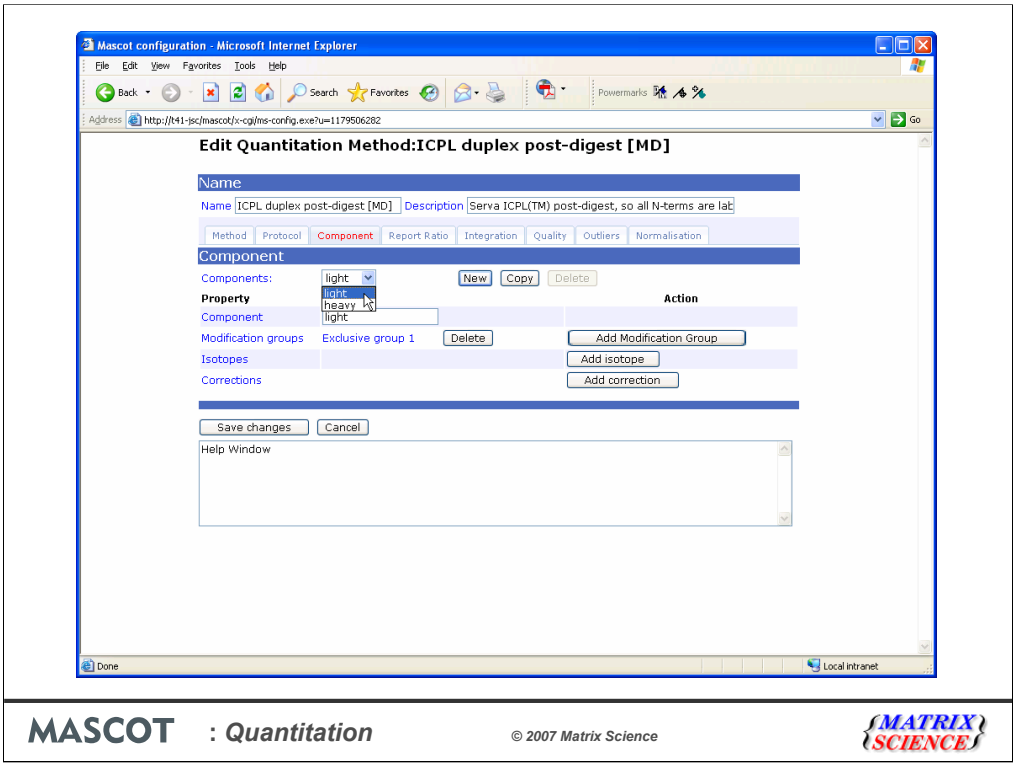
The quantitation methods are defined in a single XML configuration file

- 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



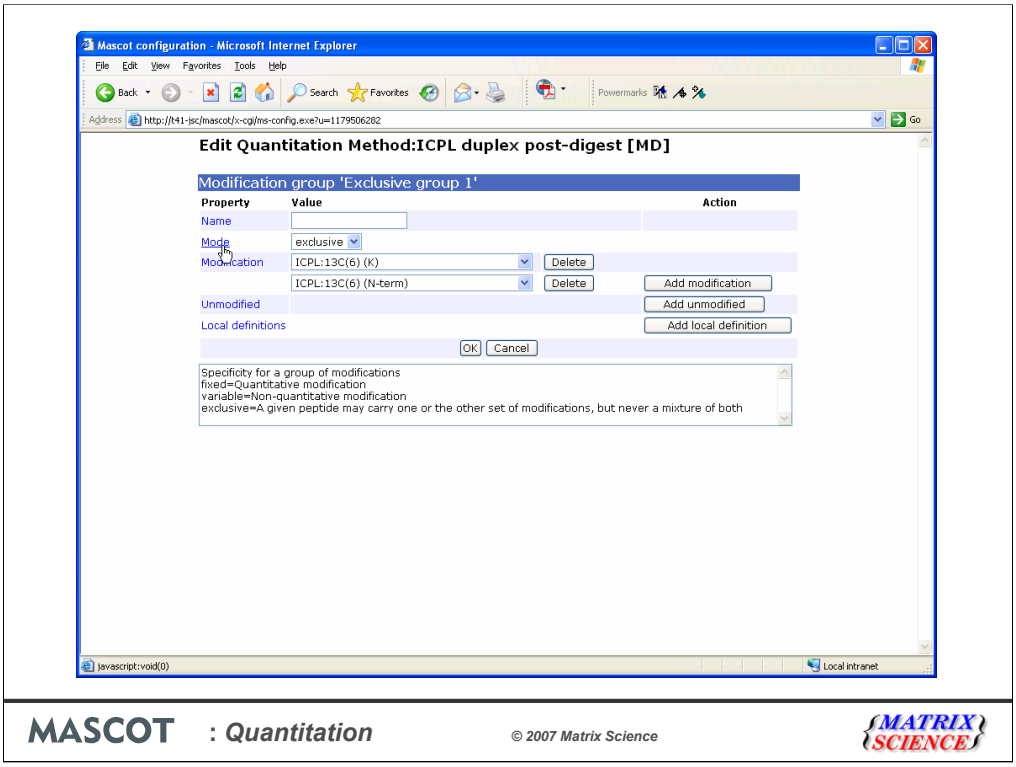
The new, browser-based Configuration Editor, provides an interface to all the Mascot configuration files. In the case of quantitation, you can edit an existing method or make a copy of it as the basis for a new method



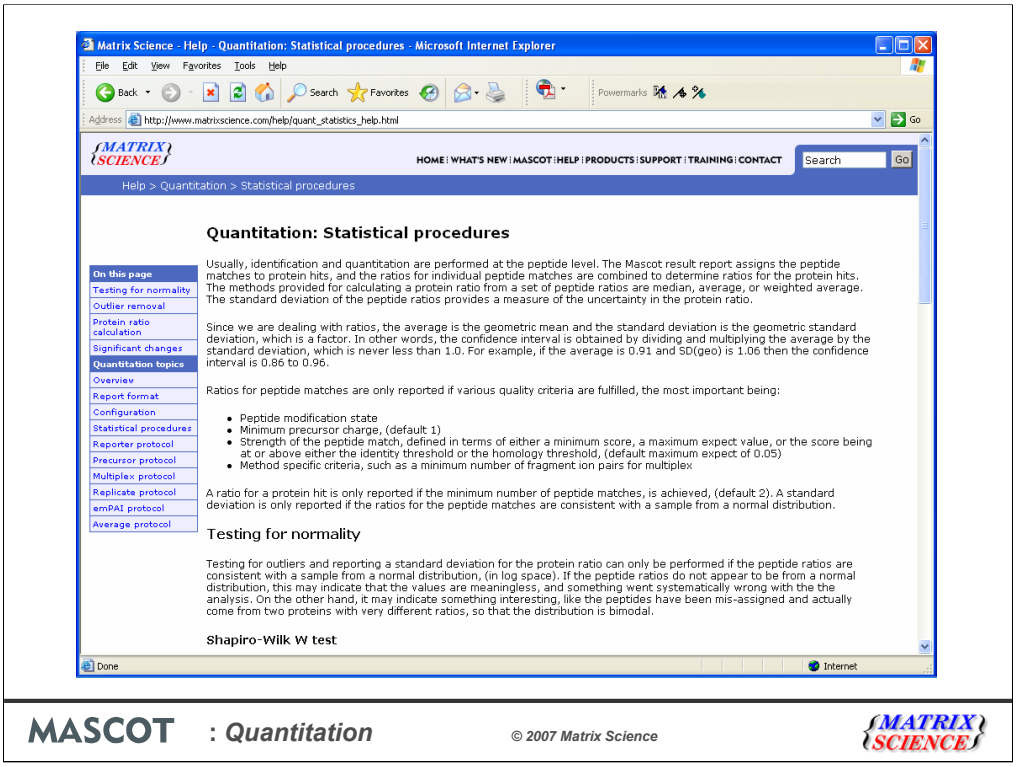
For each method, a tabbed dialog is used to navigate between property pages. In many cases, the property pages correspond to XML elements, but sometimes elements have been combined onto a single page or split across multiple pages so as to give a balanced layout.

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.

If we choose the heavy component, then click on the Modification group link

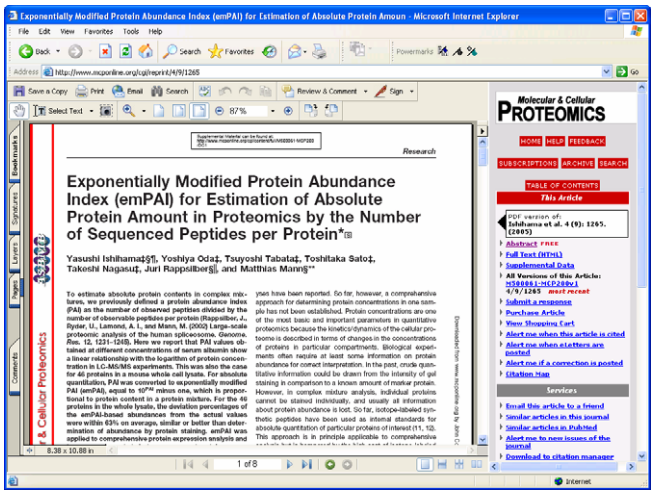


We see that the heavy component corresponds to these two modifications. In this particular method, the sample has been labelled after digestion, so the amino terminus modification is found on every peptide. As you can see from the buttons on the right, a component can be defined in terms of modified or unmodified residues or termini. A local definition is a special modification defined within the quantitation method.



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 Matrix Science* 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'**

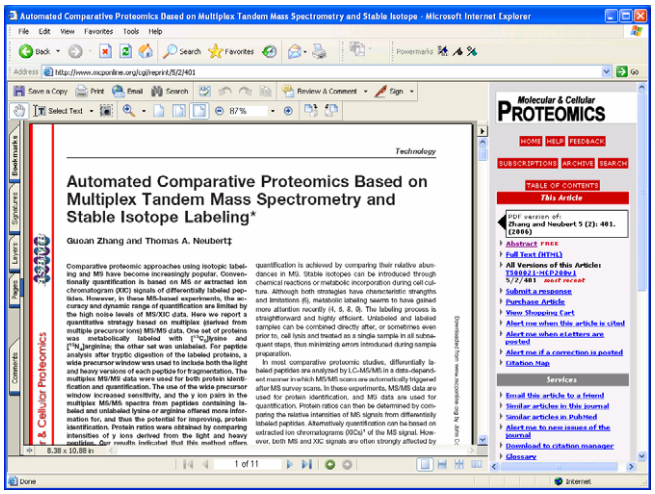
1. PPB1_HUMAN **Mass:** 58259 **Score:** 452 **Queries matched:** 17 **emPAI:** 1.04
Alkaline phosphatase, placental type precursor (EC 3.1.3.1) (PLAP-1) (Regan isozyme) - Homo sapiens
Check to include this hit in error tolerant search

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

MASCOT : Quantitation 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

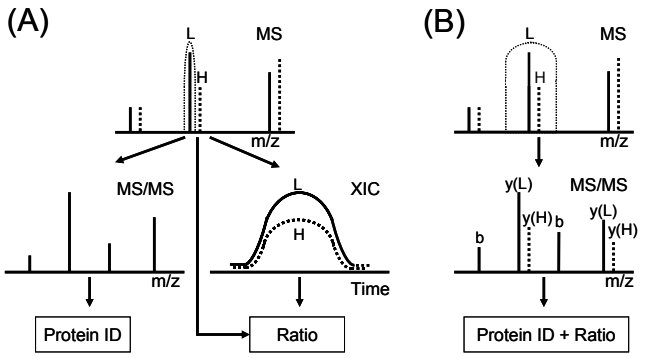
Quantitation - multiplex



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

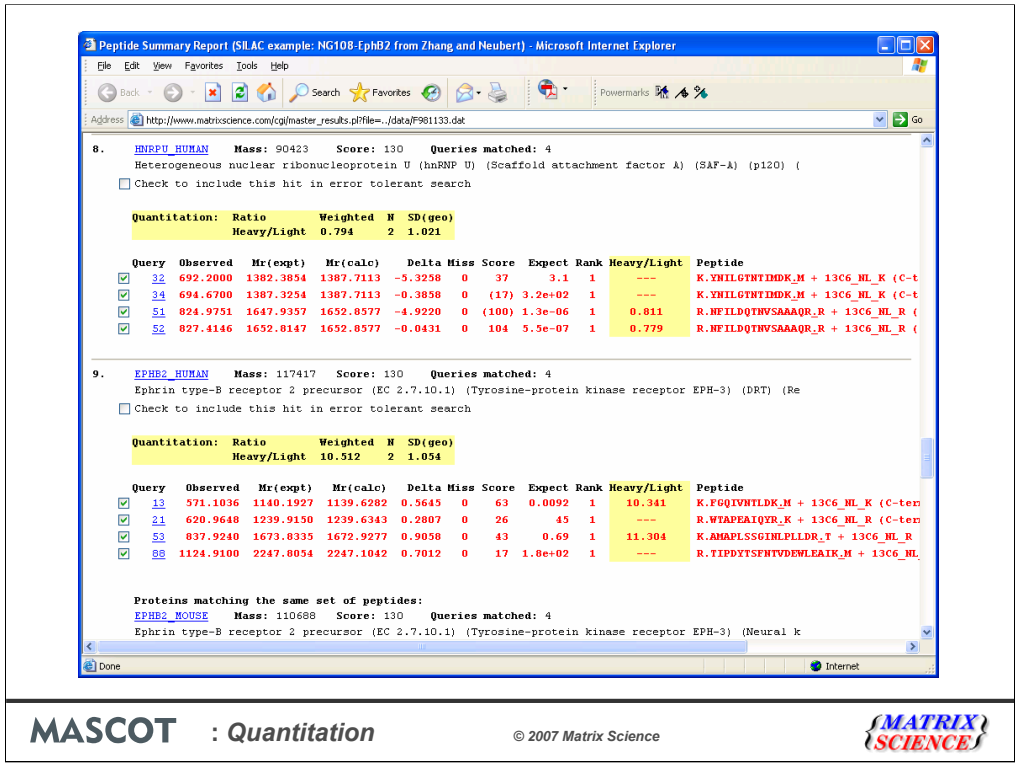
This diagram, copied from the MCP paper, illustrates how it works. On the left, we have conventional quantitation; the 'precursor protocol' in Mascot terms. This requires the precursor intensity for each component to be integrated across its elution profile. In the case of the multiplex protocol, the 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

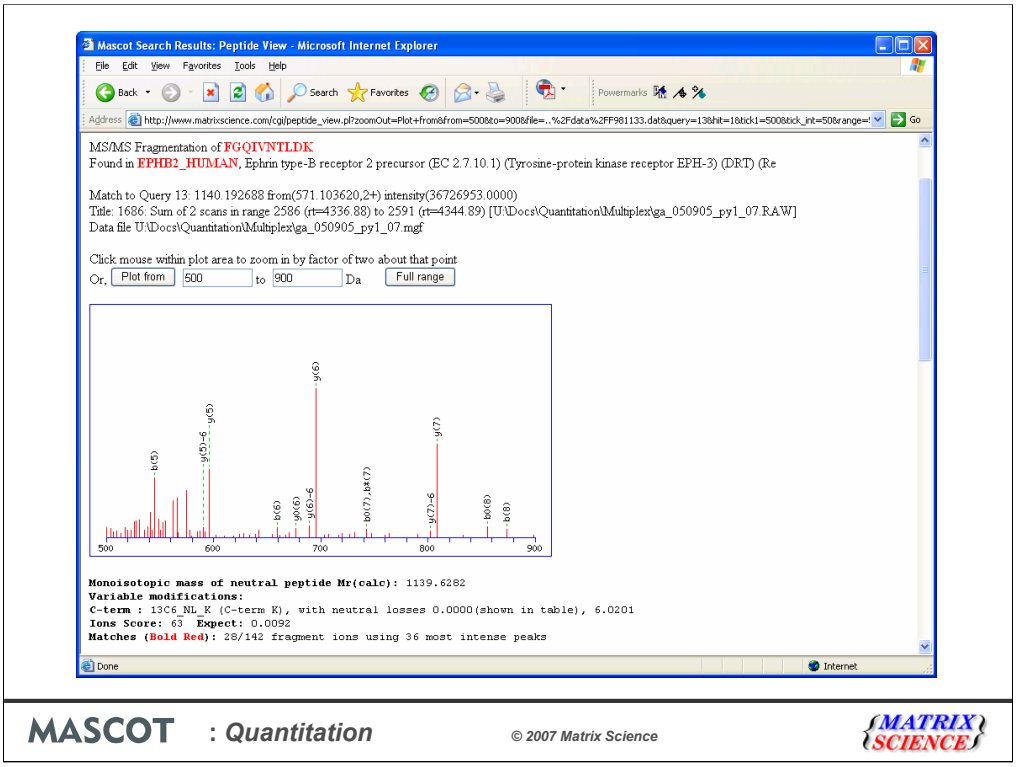
Requirements:

- 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.

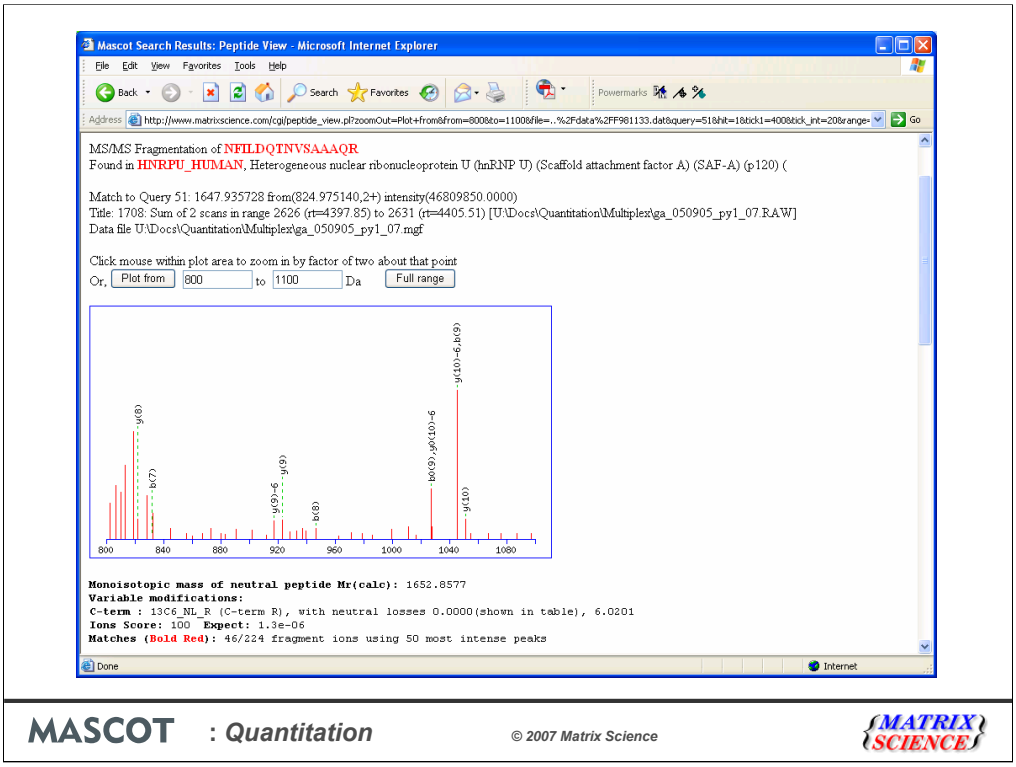
The multiplex method has the potential to give excellent precision, because each ratio is represented by multiple sequence ion pairs. On the other hand, the ratio will only be accurate if several constraints are met.



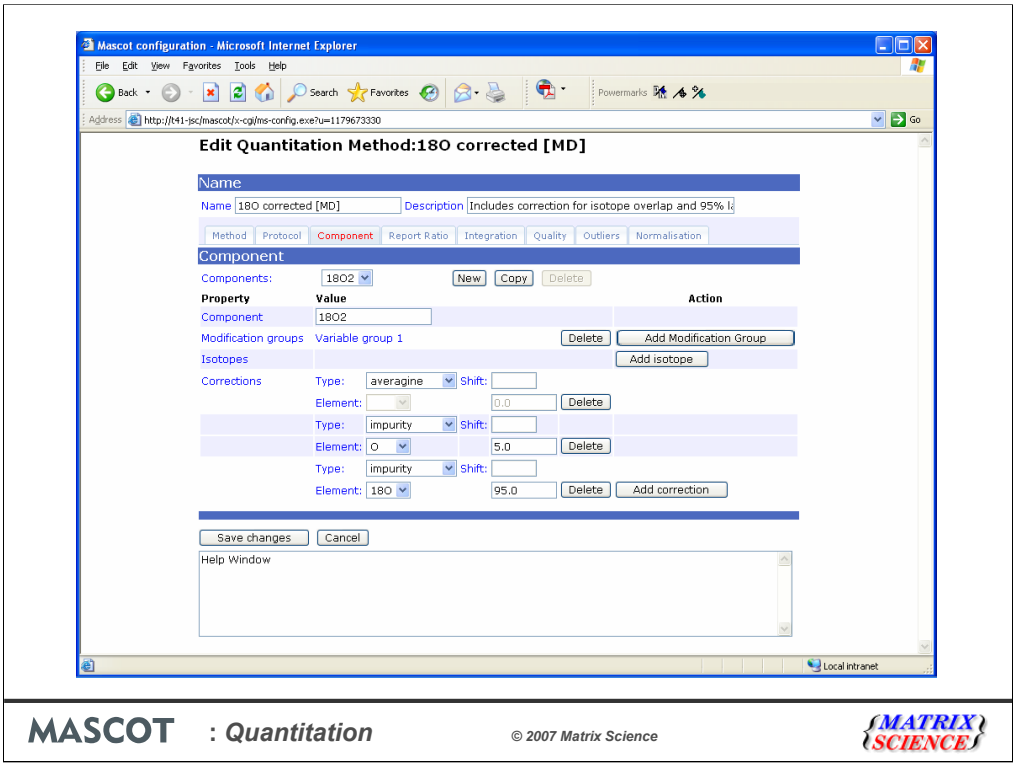
This is an example using a dataset courtesy of Zhang and Neubert. The instrument was an ion trap and the label is 13C(6) SILAC on K and R. If we look at one of the spectra from the Ephrin peptides



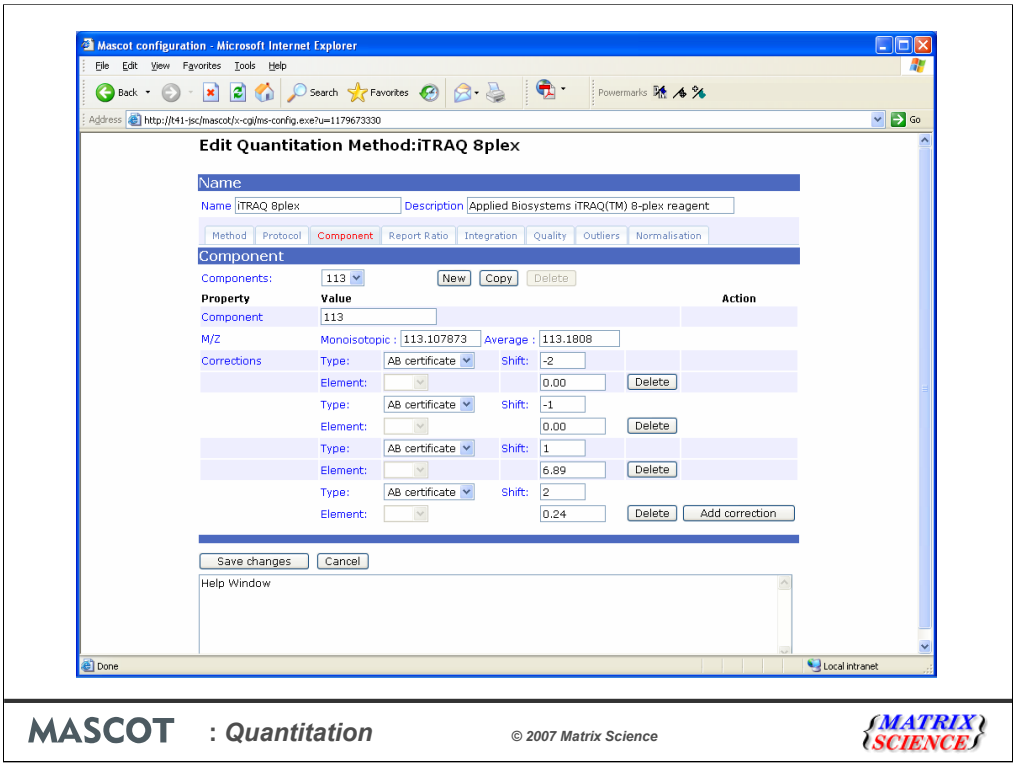
We can see that the heavy component has been strongly up-regulated



In contrast, this is a spectrum from ribonucleoprotein, which is close to 1:1. This spectrum illustrates the importance of selecting sequence ions that are not overlapped by interfering peaks. In this case, the y(10) pair has to be discarded because the light component coincides with the b(9) ion.

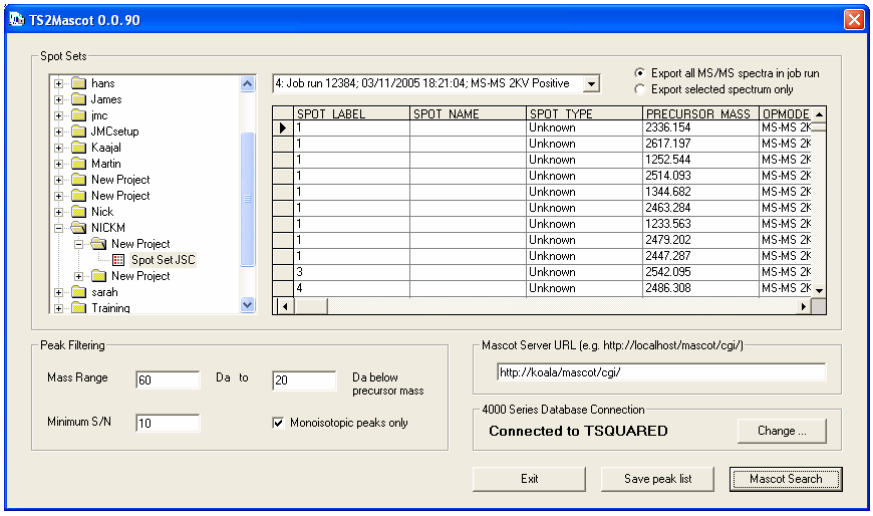


One of the complications of any type of isotope labelling is isotope impurity. It is rarely possible to get 100% enrichment. In the Mascot quantitation schema, this is described by a correction element. An ‘impurity’ correction works "downwards". That is, in this 18O method, some of the intensity of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

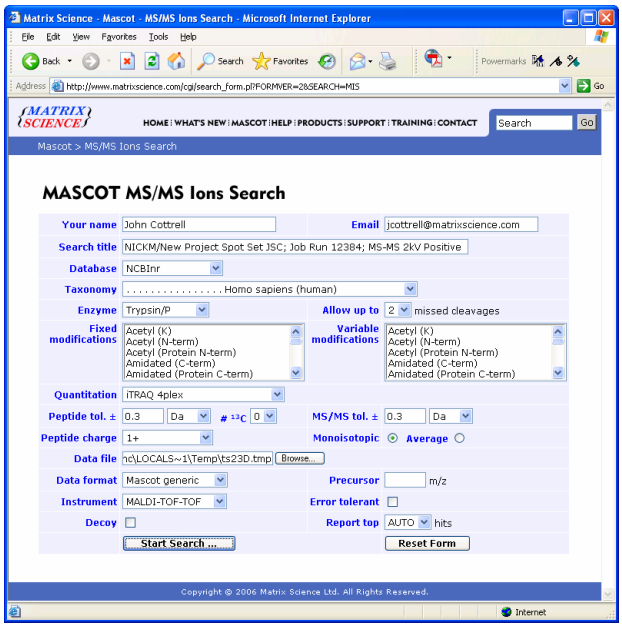
Quantitation - Reporter



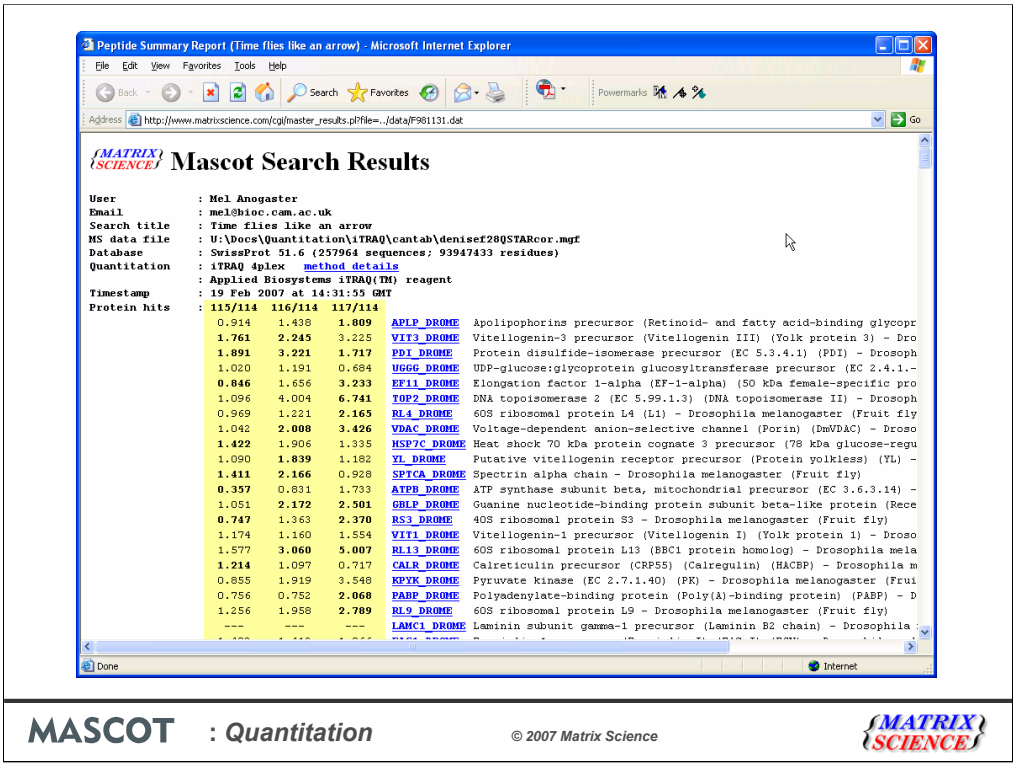
In data processing terms, the reporter protocol is one of the simplest. However, we did find that the peak list exported from the 4000 series data system or submitted to Mascot from GPS Explorer did not have the correct peak areas for the reporter ions. The numbers are different from those used within GPS Explorer for quantitation. We have had to write our own application to export a suitable 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 …

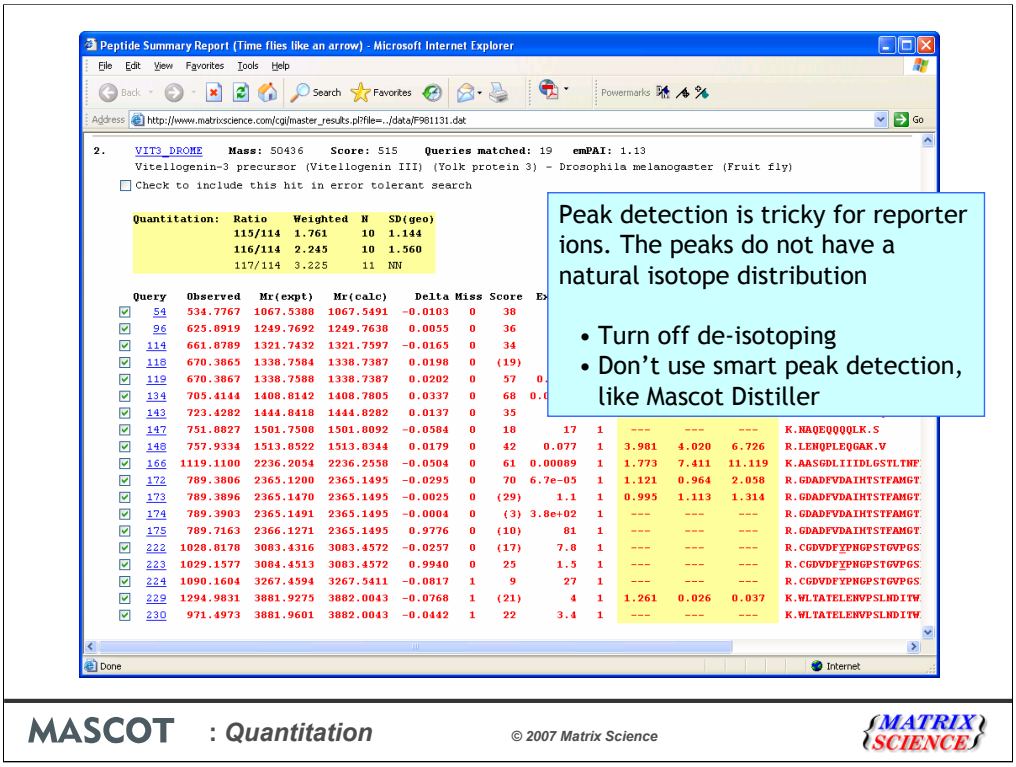
Quantitation - Reporter



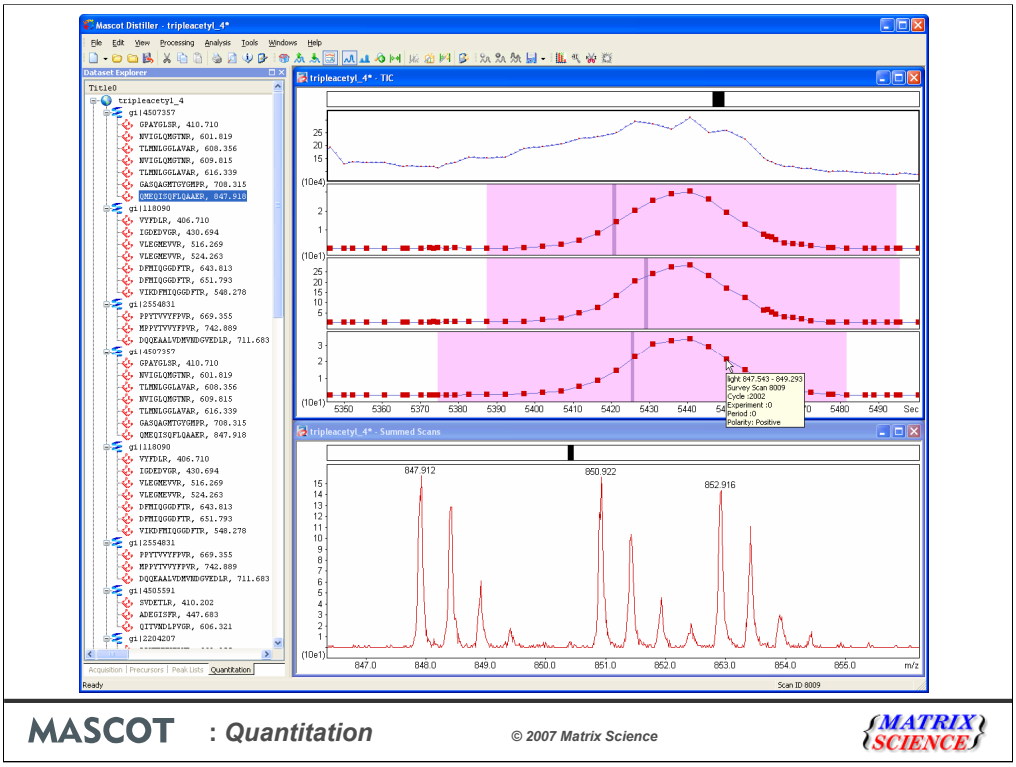
Which brings up the search form. We choose an appropriate quantitation method. We don't need to specify the iTRAQ modifications or the cysteine alkylation, because these are pre-defined in the quantitation method. Submit the search…



And back comes the report. At the top is a summary of the protein ratios. In this example, the method asks for ratios to 114, but you have total flexibility. You can edit the quantitation method to report two 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.



If you wish, you can display ratios for individual peptides. The reason there is no standard deviation for the 117/114 ratio is that it failed the normal distribution test. As you can see, the individual peptide ratios have quite a scatter. One thing you have to be very careful with is peak detection. Reporter ions do not have a natural isotope distribution, so anything that expects this, like Mascot Distiller, will not be reliable. Definitely advisable to experiment with the peak picking conditions.



Finally, I can reassure you that we are hard at work on the final piece of the jigsaw, the Quantitation Toolbox for Mascot Distiller.