

# Mascot Distiller Hot Topics

MASCOT

*MATRIX*  
*SCIENCE*

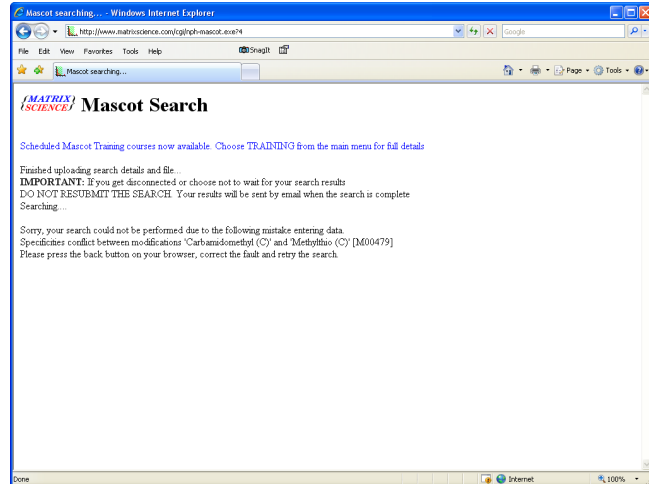
## Mascot Distiller Hot Topics

- Quantitation FAQ from 2009
- Label-free Quantitation
- Large Multi-File Projects
- Arg-Pro Conversion
- Automation

This presentation will touch on some aspects of using Mascot Distiller for quantitation. I would have called it Quantitation FAQ, but we already used that title for a presentation last year. That won't stop me taking the first few slides direct from the 2009 FAQ, because they concern three topics that keep coming up again and again.

I will then move on to discuss label-free quantitation, memory problems with very large, multi-file projects, the importance of including Arg-Pro conversion in a SILAC method to get best accuracy when arginine is being labelled, and automation of quantitation using Mascot Daemon

## 4. Why do I keep getting errors about conflicting modifications?



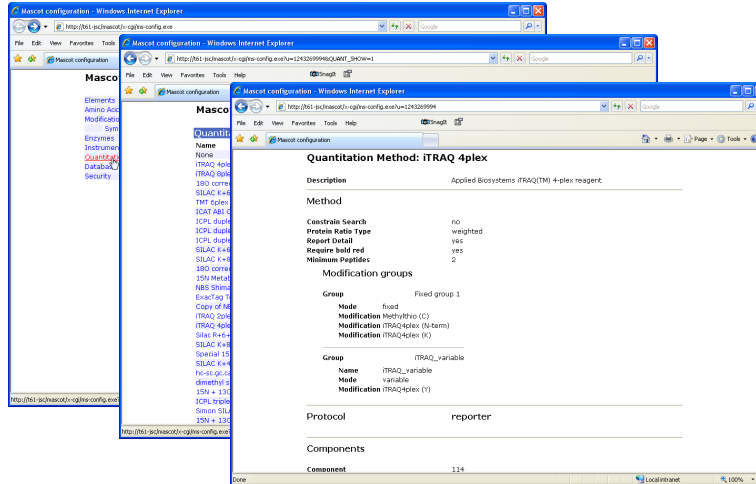
**MASCOT** : Mascot Distiller Hot Topics

© 2010 Matrix Science



First of all, the three FAQ from 2009. This question usually comes up with reference to iTRAQ. A typical error message is “Specificities conflict between modifications 'Carbamidomethyl (C)' and 'Methylthio (C)' [M00479]”

## 4. Why do I keep getting errors about conflicting modifications?



**MASCOT** : Mascot Distiller Hot Topics

© 2010 Matrix Science



To understand this, take a look at the quantitation method. You can get a pretty printed summary from the Configuration Editor. Choose Quantitation then Print for the method of interest. This is the original iTRAQ method, that shipped with Mascot 2.2. The iTRAQ modifications are defined in the quantitation method. In this particular case, this isn't strictly necessary, but for many quantitation methods the modifications have to be defined within the method because the modifications define the components. For example, a peptide with 16O at the C-terminus is light while a peptide with 18O at the C-terminus is heavy. So, the approach we took was that the isotopic labels should be defined within the method. We also felt that specifying as much as possible in the method was a good thing in terms of standard operating procedures. The original iTRAQ method was based on the Applied Biosystems procedure that used methythio (MMTS) to block the cysteines, so we also included this modification. The problem comes when someone uses a different alkylation agent, such as iodoacetamide, and chooses carbamidomethyl in the search form.

## 4. Why do I keep getting errors about conflicting modifications?

Peptide	Score	Mass	Sequence						
1096	9054	2287.6944	3286.8766	0.0170	2	0	2e+02	1	HADAARGLLMAALSTCLLPEPLGAARTK
119	691.3824	1380.7502	1380.7847	-0.0344	0	0	3.4e+02	1	PFPAGD ISLAGAR
233	1511.6445	3061.0917	3062.0250	-0.0531	2	7	1.8e+02	1	PTDGGSPGCVLTPSNITVVGKDTFAYSIYR
88	606.3040	1210.5934	1210.6293	-0.0359	1	6	4.8e+02	1	HSKAKSHTP
51	562.2309	1122.4472	1122.6267	-0.1795	0	6	2.9e+02	1	ISELPHGAR
108	877.7384	2630.1934	2630.3596	-0.1662	1	5	1.9e+02	1	ATDRLHGGHRLVLCVEEDVGR
224	1090.1404	2367.4594	2366.3897	1.0696	1	4	1.2e+02	1	DRITTTGHTMRLSTVDFGAMVNGAR
202	894.1127	2679.3163	2679.3496	-0.0334	0	4	5.3e+02	1	QSASGVEYRVALLDHVFYTTGR + iTRAQ4plex (X)
88	605.3054	1208.5962	1208.6887	-0.0924	0	4	8.3e+02	1	VVLETTLSR
187	877.3937	2629.1593	2629.2587	-0.0994	0	2	2.6e+02	1	FSPGAAALVSHFPGQARDRCR + iTRAQ4plex (X)
104	833.3796	2497.1170	2496.3104	0.8066	1	2	3.5e+02	1	TTPSYVAFHRERLIGDPAK
220	1261.0090	3782.6452	3781.6942	0.9509	0	0	1e+02	1	FSSGGSTHFLIDHMRDLSAMTFCHQIK
223	1029.1577	3084.4513	3083.3974	1.0539	0	0	7.9e+02	1	HSTYSALSSYYVYTIHQHRAEPTGAR
106	835.0141	2502.0205							
222	1028.0370	3083.4316							

**Search Parameters**

Type of search : MS/MS Ion Search  
Enzyme : Trypsin/P  
Fixed modifications : Carbamidomethyl (C), iTRAQ4plex (K), iTRAQ4plex (N-term), iTRAQ4plex (R-term), iTRAQ4plex (K)  
Mass values : Monoisotopic  
Protein Mass : Unrestricted  
Peptide Mass Tolerance : ± 0.2 Da (# 13C = 1)  
Fragment Mass Tolerance : 0.2 Da  
Max Missed Cleavages : 2  
Instrument type : ESI-QUAD-TOF  
Number of queries : 234

Mascot: <http://www.matrixscience.com/>

**MASCOT** : Mascot Distiller Hot Topics

© 2010 Matrix Science



Because it became clear that few people use MMTS, we took this out of the method when we updated it. If you haven't done so already, you should grab an updated quantitation.xml file from our web site, using the link on the Distiller support page. This fixes the MMTS problem, but it doesn't stop people using a method without looking at it. I suspect that many people have search results that look like this. The iTRAQ mods are there twice, because they have been chosen in the search form even though they are already in the method. This isn't an error, because there is no conflict, but it does indicate that we haven't got the message across about embedding the important search parameters in the method. Hopefully, this FAQ will help clarify that the modifications that are essential to the method should go in the method while those that are 'optional' and will vary from experiment to experiment, such as oxidation or acetylation, should be selected in the search form.

## 7. Why do I get no results for certain peptide matches?

Query	Observed	Mr(calc)	Mr(calc)	Delta Miss Score	Expect	Rank	115/114	116/114	117/114	Peptide		
53	534.7767	1067.5288	1067.5491	-0.0103	0	38	0.2	1	1.962	15.529	13.409	K.FQGFQR.R
26	625.0919	1249.7692	1249.7638	0.0053	0	36	0.35	1	2.148	2.180	3.490	K.LIGDVFQR.Y
113	661.0789	1321.7432	1321.7597	-0.0165	0	34	0.52	1	1.929	1.519	2.946	K.YHQQLQK.W
110	670.3065	1338.7584	1338.7387	0.0198	0	(19)	17	1	---	---	---	K.LHVVYTAK.A
119	670.3067	1338.7589	1338.7307	0.0282	0	57	0.0023	1	1.912	1.956	3.475	K.LHVVYTAK.A
124	705.4144	1408.8142	1408.7805	0.0337	0	68	0.00018	1	1.027	5.873	10.012	R.GDVFVAINTSTFAMGT
143	723.4282	1444.8418	1444.8282	0.0137	0	35	0.4	1	1.174	0.851	1.783	R.RFDFVANSK.R.Q
147	751.0827	1501.7588	1501.8092	-0.0504	0	18	17	1	---	---	---	K.MAQFQQQLK.S
148	757.9334	1513.8522	1513.8344	0.0179	0	42	0.077	1	3.981	4.020	6.726	R.LEHQLEQAK.V
166	1119.1100	2236.2054	2236.2558	-0.0504	0	61	0.00089	1	1.773	7.411	11.119	K.AASDLIIDLQSLTWF
172	789.3096	2365.1470	2365.1495	-0.0025	0	(29)	1.1	1	0.995	1.113	1.314	R.GDVFVAINTSTFAMGT
174	789.3903	2365.1491	2365.1495	-0.0004	0	(3)	3.8e+02	1	---	---	---	R.GDVFVAINTSTFAMGT
175	789.7163	2366.1271	2365.1495	0.9776	0	(10)	81	1	---	---	---	R.GDVFVAINTSTFAMGT
222	1028.0178	3083.4316	3083.4572	-0.0257	0	(17)	7.8	1	---	---	---	R.CGDVDFYFNGPSTGVP
223	1029.1377	3084.4513	3083.4572	0.9940	0	23	1.5	1	---	---	---	R.CGDVDFYFNGPSTGVP
224	1090.1604	3267.4594	3267.5411	-0.0817	1	9	27	1	---	---	---	R.CGDVDFYFNGPSTGVP
229	1294.9831	3881.9275	3882.0043	-0.0768	1	(21)	4	1	1.261	0.026	0.037	K.WLATELVVPSLSDITW
230	971.4973	3881.9681	3882.0043	-0.0442	1	22	3.4	1	---	---	---	K.WLATELVVPSLSDITW

**MASCOT** : Mascot Distiller Hot Topics

© 2010 Matrix Science



The second repeated FAQ: A Mascot quantitation method includes certain rules and quality thresholds. One of the most important is the Mascot score. If you aren't confident of the peptide match, then you can't be confident of the protein assignment for that match and wouldn't want to include it in any calculations. A cut-off, which is often used in the default methods, is that the score should be above the homology threshold. So, first thing to check is that the ions score for the match is above threshold. Here, this is the reason for most of the missing ratios. It isn't the explanation for the match to CGDVDFYFNGPSTGVPGENVIEAVAR with a score of 25. This has been excluded by a rule in the method, that any match with iTRAQ on tyrosine should be ignored. This is because the reaction with tyrosine is slower than with lysine and the amino terminus, which means that the ratio is likely to be skewed.

## 7. Why do I get no results for certain peptide matches?

The screenshot shows the 'Quantitation Method: SILAC R-6 R-10 [MD]' dialog box. The 'Allow mass time match' checkbox is checked. Below the dialog box, there is a note: 'Allow mass time match. If false, we only report a ratio if we have peptide matches to both components. Also, if false, then all\_charge\_states treated as false, even if true in method. Almost always set true in practice.'

Method	Value
Protein Ratio Type	median
Protein Score	msdpt
Report Detail	<input checked="" type="checkbox"/>
Show subsets	0.20
Require Bolded	<input checked="" type="checkbox"/>
Minimum Peptides	2
Significance Threshold	0.05

Protocol	Value
precursor	precursor
Allow mass time match	<input checked="" type="checkbox"/>

Components	Value
Report Ratios	<input checked="" type="checkbox"/>

Integration	Value
Integration Method	trapezium
Single Ratio	<input type="checkbox"/>
Integration Source	survey
Allow elution time shift	<input type="checkbox"/>
Elution Time Delta	10 seconds
Std. Err. Threshold	0.2
Correlation Threshold	0.9
XIC Threshold	0.1
All charge states	<input type="checkbox"/>
All charge state threshold	0.20
Max XIC width	250
XIC Smooth	3

Quality	Value
Outlier Removal	<input checked="" type="checkbox"/>
Normalisation	<input checked="" type="checkbox"/>

**MASCOT** : Mascot Distiller Hot Topics

© 2010 Matrix Science

**MATRIX**  
**SCIENCE**

If you have less peptide matches than expected, check that 'Allow mass time match' is true. If this is false, then you must have peptide matches for all components before a ratio can be reported. If true, you only need a peptide match to one component. Its partner or partners will be identified from their molecular mass and elution time. This is almost always what is wanted.

## 10. Should I choose normalisation? If so, which type is best?

### When to use 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

The final repeated FAQ concerns normalisation. The correct time to use normalisation is when you have two or more samples, each of which is supposed to contain the same total amount of a reasonably complex mixture of proteins. Some proteins may be up-regulated, others down-regulated, but the majority are unchanged. In such a case, the mean ratio is expected to be close to unity. If it is not unity, this probably reflects systematic errors, such as failure to fully label or digest a sample, or maybe the sample aliquots were not identical. By forcing the mean ratio to unity, we can eliminate these systematic errors.

That is, normalisation forces the average or median peptide ratio to be 1. If the average or median ratio is supposed to be 1, this is the smart thing to do. In other cases, it is absolutely the wrong thing to do. For example:

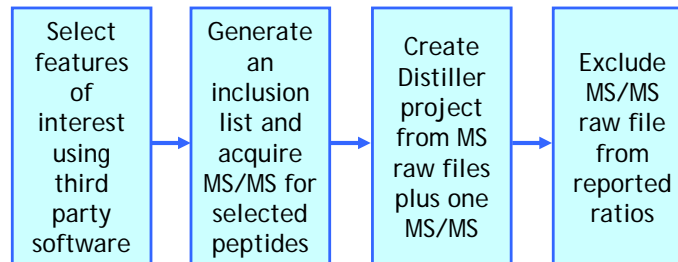
- If you are analysing a standard, where the ratio is supposed to be 3:1, you wouldn't want to force it to be 1:1, would you?
- If you start with equal numbers of cells and want to see the effect of starvation on one line then normalisation would be incorrect because you expect less protein from starved cells
- If you are isolating a sub-set of proteins by affinity methods then normalisation is not appropriate

Whether you get better results using average or median depends on various factors. The safest way to choose is to run some standards (e.g. similar samples spiked with varying amounts of a known protein) and see which gives the best accuracy and precision.



## Label-free Quantitation

- Mascot Server must be 2.2.06 or later
- Each raw file is a component
- Search result driven, not feature driven
- But ...



**MASCOT** : Mascot Distiller Hot Topics

© 2010 Matrix Science

**MATRIX**  
**SCIENCE**

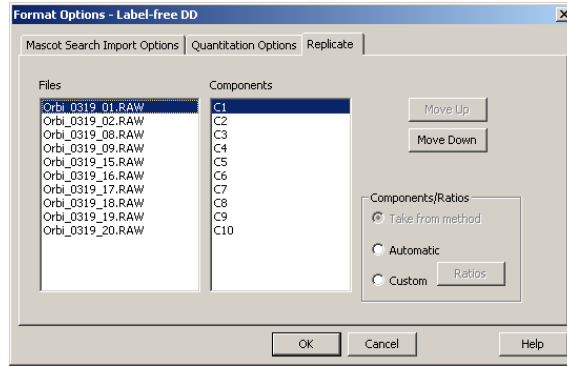
Label-free, or Replicate, was the final quantitation protocol to be added to Distiller, in version 2.3. It requires Mascot Server 2.2.06 or later.

This is quantitation based on the relative intensities of extracted ion chromatograms (XICs) for precursors in multiple analyses, aligned using mass and elution time. Each raw file is a component and the chromatography needs to be highly reproducible.

Within Distiller, quantitation is driven by the search results. That is, Distiller takes the list of peptides returned by the Mascot search and looks for the precursors in each of the survey scans. In most cases, the majority of proteins are unchanged from sample to sample. Only a small number are significantly up or down regulated. The alternative workflow is to align multiple files of MS data, identify features that are up or down regulated, generate an inclusion list for a mass spectrometer to obtain MS/MS data for these peptides, import the resulting peptide matches, then perform quantitation on just the features of interest.

Distiller doesn't provide such an interface but, if you have some third party software that can be used to generate an inclusion list, you can easily create a Distiller project containing the MS/MS data in a separate file. Distiller will happily use the MS/MS to locate the set of features to be quantified across all the MS files. Just omit the MS/MS file from the ratios to be reported.

## Label-free Quantitation

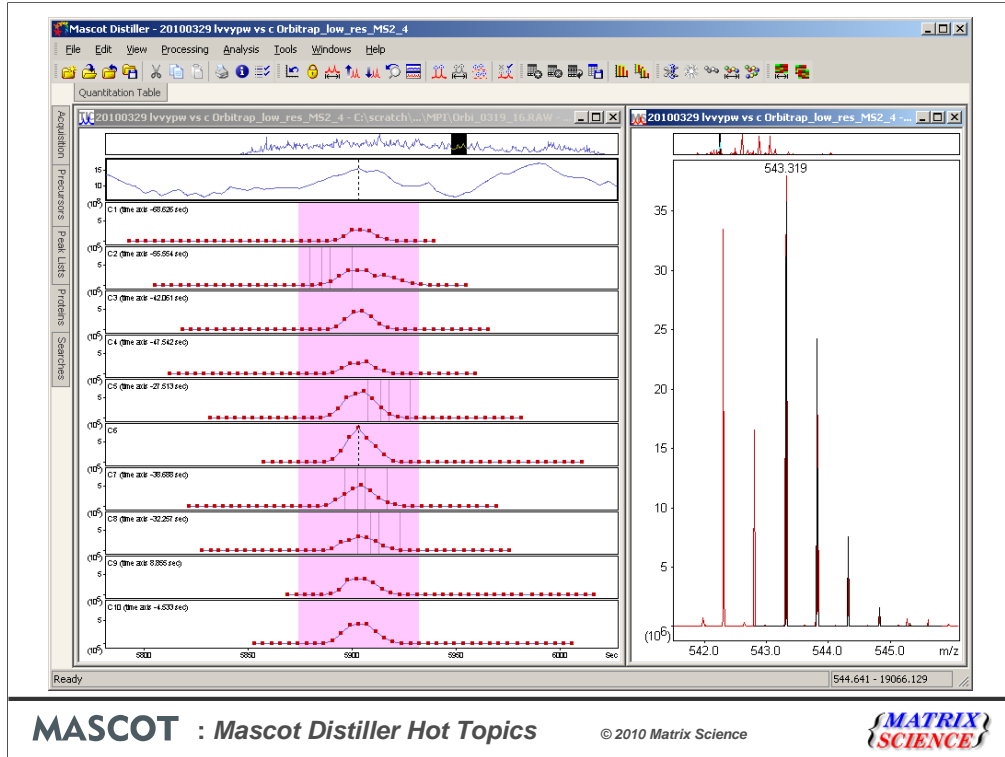


**MASCOT** : Mascot Distiller Hot Topics

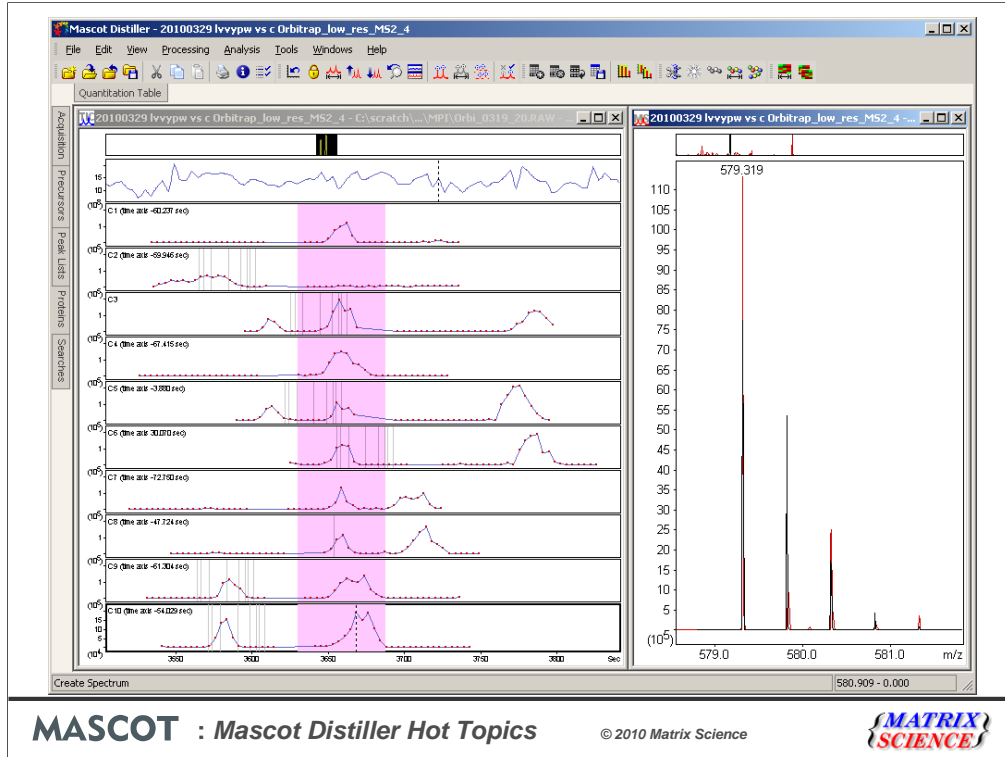
© 2010 Matrix Science

**MATRIX**  
**SCIENCE**

A replicate project is always a multi-file project, with one file for each component. You can define 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.

## Large, Multi-file Projects

### Current release of Distiller (2.3.2) cannot handle large, multi-file projects

- Problem typically seen when total size of all raw files > 5 GB
- 64-bit Windows with plenty of free RAM helps, but is not a fix
- Recommend saving large projects after time-consuming steps.

Quite often, a label-free experiment will comprise a large number of raw files. Unfortunately, the current release of Distiller runs into problems with multi-file projects when the total size of all the raw files is several GB. It isn't a precise threshold – depends on size of peak list, number of peptide matches, and many other factors.

Running under 64-bit Windows with plenty of RAM helps, but is not a fix. If working with a large set of raw files, its a good idea to save the project after each time consuming step, so as not to lose the results if Distiller should crash.

Distiller is a 32-bit application and it is trying to do too much in memory. We are working hard on a fix and hope to have something later this summer

## Arg-Pro Conversion

**Article** Prev.

**Properties of  $^{13}\text{C}$ -Substituted Arginine in Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC)**


Shao-En Ong, Irina Kratchmarova, and Matthias Mann\*

Center for Experimental Bioinformatics (CEBI),  
Department of Biochemistry and Molecular Biology,  
University of Southern Denmark, Campusvej 55,  
DK-5230 Odense M, Denmark

*Journal of Proteome Research*, 2003, 2 (2), pp 173-181  
DOI: 10.1021/pr0255708  
Publication Date (Web): December 13, 2002  
Copyright © 2003 American Chemical Society

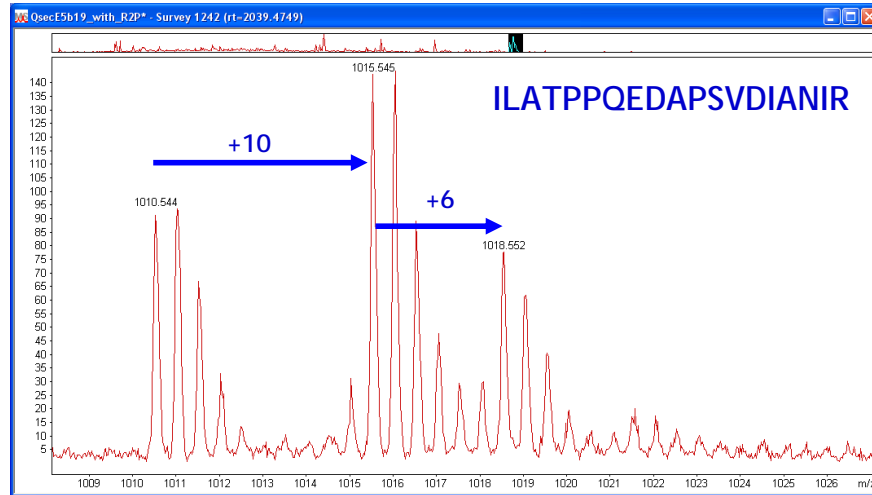
**Abstract**

- Full Text HTML
- Hi-Res PDF [206 KB]
- PDF w/ Links [100 KB]

**MASCOT** : Mascot Distiller Hot Topics © 2010 Matrix Science 

Ong reported that cells grown in media containing labelled arginine could yield peptides containing labelled proline. To obtain an accurate ratio, it becomes necessary to account for the label distributed across these additional peaks.

## Arg-Pro Conversion



**MASCOT** : Mascot Distiller Hot Topics

© 2010 Matrix Science

**MATRIX**  
**SCIENCE**

This can be handled in a very general way in Distiller by defining a satellite modification group. Distiller sums the intensity of any corresponding satellite peaks into the heaviest component. (This is only practical for the heaviest component. If the method includes multiple labelled components, it is not possible to de-convolute the satellite peaks for intermediate components.)

The proline label is not identical to the arginine label. For example, Arg labelled with  $^{13}\text{C}(6)$  converts to Pro labelled with  $^{13}\text{C}(5)$ , while  $^{13}\text{C}(6)^{15}\text{N}(4)$  converts to  $^{13}\text{C}(5)^{15}\text{N}(1)$ .

## Arg-Pro Conversion

The screenshot shows a web browser window titled "Mascot configuration - Windows Internet Explorer". The address bar shows a URL starting with "http://rs1-isc/mascot/...". The page content is titled "Edit Quantitation Method: SILAC K+6 R+10 Arg-Pro [MD]".

Under the heading "Modification group 'satellite'", there is a table with the following structure:

Property	Value	Action
Name	satellite	
Mode	variable	
Required	<input type="checkbox"/>	
Modification		<input type="button" value="Add modification"/>
Unmodified		<input type="button" value="Add unmodified"/>
Local definitions		<input type="button" value="Add local definition"/>

Below the table, there is a form for defining a new modification:

Name: ArgPro

Composition:

Specificity: Site: P Position: Anywhere

Help Window:

**MASCOT** : Mascot Distiller Hot Topics

© 2010 Matrix Science

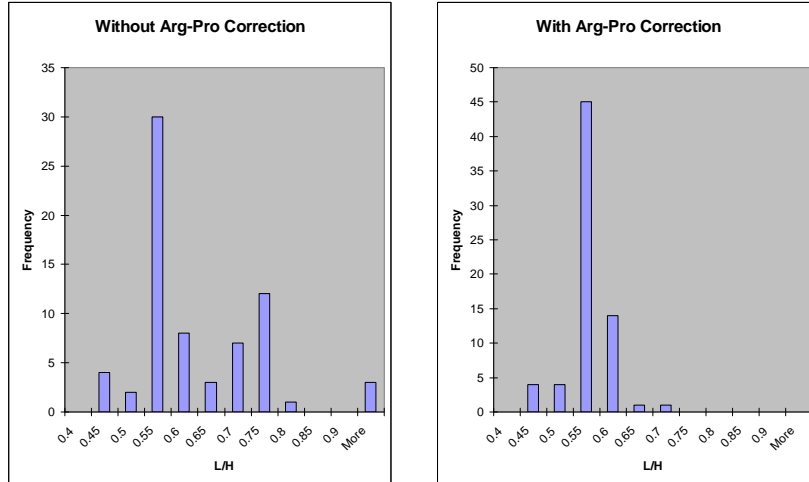


To add this to a new method, from the Component tab, select the heavy component and choose Add Satellite Group.

Usually, it will be easiest to define this as a local modification definition, within the quantitation method, as shown here



## Arg-Pro Conversion



**MASCOT** : Mascot Distiller Hot Topics

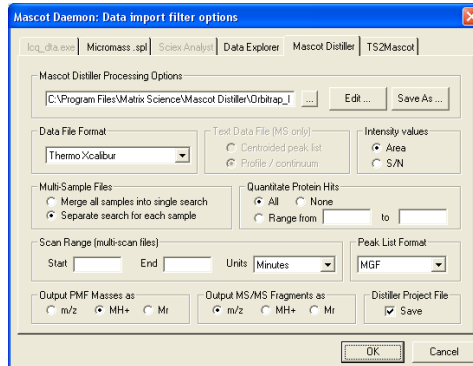
© 2010 Matrix Science

**MATRIX**  
**SCIENCE**

This histogram shows the improvement for a single protein. This is Orbitrap data and the labelling was K+6 and R+6. We've made a histogram of all the peptide ratios. Without a correction, we have a bi-modal distribution. The proline containing peptides give a ratio that is too high because part of the heavy intensity has been shifted to the peptide with heavy proline. When we correct for Arg-Pro conversion, the second peak pretty much disappears and we have a much narrower distribution

## Automation

All steps from peak picking to quantitation can be automated using Mascot Daemon 2.3



**MASCOT** : Mascot Distiller Hot Topics

© 2010 Matrix Science



Finally, just to mention that Mascot Daemon release 2.3 is capable of automating all stages of a quantitation experiment. You choose to quantitate all the proteins or just a range, such as hit 1 or hits 1 to 10. When quantitation is complete, Daemon saves the the results to the Distiller project file.

One limitation: Daemon can only automate single file Distiller projects, not multi-file projects. But, if you think about it, you wouldn't gain much by using Daemon for a multi-file project. Easier just to select the set of files in Distiller.