

Improved variable modification handling in Mascot 2.7

Patrick Emery
Matrix Science Ltd

MASCOT : *Modifications*

© 2020 Matrix Science



Types of Modifications

- **Post-translational**
 - E.g. Phosphorylation, acetylation
- **Artefacts**
 - E.g. Oxidation, acetylation
- **Derivatisation**
 - E.g. Alkylation of cysteine, SILAC

MASCOT : *Modifications*

© 2020 Matrix Science



Modifications are an important topic in database searching. The main focus of a study may be to characterise post translational modifications with biological significance – for example, studying phosphorylation patterns

In other cases, the modification may not be of interest in itself, but may be required to get a match during the database search. Oxidation during sample preparation is a common example of this.

Some modifications are the result of deliberate derivatisation as part of the experiment – for example, using isotopic labels for quantitation.

During the database search, modifications can be handled in two ways – Fixed modifications are where ever instance of the target residue are modified. These are simple to handle because it is effectively just a change in the mass of the residue. The more complex case is variable modifications, where just a limited number of the residues are modified. In this presentation we'll be looking at some very important changes in how Mascot permutes the possible arrangements for variable modifications during a database search.

New features in Mascot 2.7

Other new features

Fragment charge

If fragment charge is present as a 3rd column in an MS/MS peak list, fragments can be matched as if singly charged, which will usually improve the score.

Better filtering of modifications with complementary deltas in an error tolerant searches

For example, a match that gets a fractionally better score by adding 16 Da at one residue and dropping oxidation of a nearby Met can now be suppressed.

Combined search results in Mascot Daemon

Multiple searches, such as fractions, can be selected in Daemon and a merged summary report opened in a web browser

MGF export outputs spectra in the original order

An option has been added to export the peak list in its original order, rather than sorted by precursor M_r.

Improved limits on variable modifications

Better site localisation by introducing separate caps on the number of distinct modifications in a PSM, the number of modified sites in a PSM, and the total number of arrangements to be tested.

The quantitation code in Mascot Distiller and Mascot Server has been moved to the Mascot Parser library, making it available to user programs and scripts.

MASCOT : Modifications

© 2020 Matrix Science

MATRIX
SCIENCE

When we released Mascot 2.7 in January this year, tucked away under ‘Other new features’ was this:

“Improved limits on variable modifications:

Better site localisation by introducing separate caps on the number of distinct modifications in a PSM, the number of modified sites in a PSM, and the total number of arrangements to be tested.”

So, what exactly do we mean by this, and why is this a significant improvement to the way Mascot can handle variable modifications?

Variable Modification Permutation

KKKSTKKSTKSKSK

Acetyl (K), Phospho (ST)

- 1 x Acetyl (K) : 8 arrangements
- 1 x Phospho (ST) : 6 arrangements
- 2 x Acetyl (K) : 28 arrangements
- 2 x Acetyl (K) + 1 x Phospho (ST) : 168 arrangements
- ... and so on ...

MASCOT : Modifications

© 2020 Matrix Science

MATRIX
SCIENCE

First let's define what it is the software is having to do when looking for modified sites.

Let's consider this slightly unusual peptide comprised of 8 Lysine, 4 serine and 2 threonine residues as an example. If we carried out a search with Lysine acetylation and Phospho Serine & Threonine:

If the precursor mass and tolerance allowed for a single Lysine acetylation, there are just 8 possible arrangements of this.

Likewise, if a single phosphorylation was possible, there are just 6 possible arrangements. However, the number of possible arrangements increases rapidly as we consider more modifications and modifiable sites.

If we need two lysine residues to be acetylated to match the precursor, we now have 28 possible arrangements.

Add in a single phosphorylation and we now have the possible 28 acetylation sites combined with 6 possible phosphorylation sites for a grand total of 168 possible arrangements.

This is the so call combinatorial explosion, and it's one reason why some searches with a large number of frequent modifications can take a long time

Modification permutation in Mascot 2.6 or earlier

- No upper limit on no. modified sites
- Permutation has built in limits
- Number of arrangements < 16
 - All are tried
- Greater than 16 'sliding window' used
 - Testing all possibilities would be too slow

Here's how variable modification permutation works in Mascot 2.6 or earlier.

There is no upper limit on the number of modified sites per peptide. However, permutation of the modifications options does have built in limits.

If the maximum number of arrangements for a peptide is less than 16, then all possible permutations are tested for matching by Mascot.

However, if there are more than 16 arrangements, then a second approach is automatically used, where a sliding window is applied to the peptide. This is to prevent the search from getting too slow and taking too long.

Problems with 2.6 approach

- **Less than 16 possible permutations:**
 - No issues, all possibilities tested
- **More than 16 possible permutations:**
 - Tends to cluster modifications on adjacent modifiable sites
 - Often stops before 16 different permutations tested

In general, the approach works well. However, it isn't without its limitations.

If a peptide has less than 16 possible variable modification permutations, then there are no issues as all possibilities are tested. It's in the cases where the peptide has more than 16 possible permutations that issues can arise.

The sliding window method tends to cluster modifications on adjacent modifiable sites, and it will often stop before 16 different permutations have been tested.

Modification iterator in Mascot 2.7

- **Single, consistent, permutation method**
 - No switching between methods
- **Controlled by 3 user definable parameters:**
 - **MaxPepNumVarMods**
 - Max no. of different variable modifications per peptide
 - **MaxPepNumModifiedSites**
 - Max no. of modified residues per peptide
 - **MaxPepModArrangements**
 - Max no. of arrangements of an individual varmod composition

MASCOT : *Modifications*

© 2020 Matrix Science



In Mascot 2.7 we've taken a different approach. We use a single, consistent permutation method – there's no switching between different methods. The new permutation iterator samples arrangements using a uniformly random scheme. The operation of this is controlled by 3 user definable settings.

MaxPepNumVarMods – this specifies the maximum number of different variable modifications which can be applied to a peptide

MaxPepNumModifiedSites – this specifies the maximum number of residues which can be modified on a peptide

MaxPepModArrangements – this specifies the maximum number of arrangements of an individual varmod composition to test

Modification iteration in Mascot 2.7

- **Defaults chosen to give similar speed and depth of search to Mascot 2.6 or earlier**
 - MaxPepNumVarMods 3
 - MaxPepNumModifiedSites 5
 - MaxPepModArrangements 64
- **Two main cases for changing the defaults**
 - Decrease limits to reduce search time
 - Increase limits to improve site analysis, or if you're looking at a highly modified protein

MASCOT : *Modifications*

© 2020 Matrix Science



Mascot 2.7 ships with the following default values for these parameters:

MaxPepNumVarMods 3, MaxPepNumModifiedSites 5 and MaxPepModArrangements 64 – these have been chosen to give a similar speed and depth of search to Mascot 2.6 or earlier. So for most searches with variable modifications, you won't see major differences in the results if you repeat an old search on Mascot 2.7. Assuming you're using the same database release of course.

There are two main cases where you might want to change these defaults. Decreasing any of these values will reduce the search space, as fewer arrangements will be tested. This will decrease the search time – so if you're looking at a sample which is not highly modified and where definitive site analysis is not the aim of the study, you may wish to decrease some of these values.

However, if site analysis is important, or if you're looking at a highly modified protein such as Histone, then you may need to increase these limits in order to gain accurate modification localisation results.

We'll take a look at examples of each of these cases later in the presentation.

Modification iteration in Mascot 2.7

- Values can be set globally in Mascot.dat

| | |
|------------------------|----|
| MaxPepNumVarMods | 3 |
| MaxPepNumModifiedSites | 5 |
| MaxPepModArrangements | 64 |

- ‘Locally’ in MGF header

```
OPTION_MaxPepNumVarMods=7  
OPTION_MaxPepNumModifiedSites=8  
OPTION_MaxPepModArrangements=256
```

If you need to change these settings, then you can set the values globally via the Configuration Options under Mascot Configuration on your Mascot server.

However, it can be difficult to set limits which are good for all experiments – if you have a middle or top down experiment alongside a standard bottom up experiment, you may well need very different settings for searching data from each experiment. In those cases you can specify the parameters ‘locally’ by embedding them at the head of the MGF peaklists file – the embedded parameters will override the defaults specified by your Mascot configuration.

Mascot Daemon tip

```
1 COM=<taskname> (<parameters>), submitted from Daemon on <localhost>
2 SEG=
3 TOL=50
4 TOLU=ppm
5 ITOL=10
6 ITOLU=ppm
7 PFA=0
8 DB=Histone_H4_variants
9 MODS=
10 MASS=Monoisotopic
11 CLE=Asp-N
12 SEARCH=MIS
13 USERNAME=<mascot_user_full_name>
14 USEREMAIL=<mascot_user_email>
15 CHARGE=2+ and 3+
16 REPORT= AUTO
17 FORMAT=Mascot generic
18 IT_MODS=Acetyl (K)@Acetyl (Protein N-term)@Dimethyl (K)@Methyl (K)@Methyl (R)@Phospho (ST)@Trimethyl (K)
19 TAXONOMY=All entries
20 INSTRUMENT=ETD-TRAP-1Plus
21 PEP_ISOPE_ERROR=2
22 OPTION_MaxPepModArrangements=64
23 OPTION_MaxPepNumModifiedSites=11
24 OPTION_MaxPepNumVarMods=7
```

MASCOT : *Modifications*

© 2020 Matrix Science



If you're using Mascot Daemon and using a data import filter like Mascot Distiller to process raw files into peaklists, then you can add additional options like these to the .par parameter files – just add them at the end like I've done in this example. This will allow you to add the parameters when you don't have immediate access to the peaklists which will be submitted, or if you're searching peaklists in a format other than MGF.

Example 1: Speeding up an error tolerant search

- Subset mouse label free sample (PXD013086)
- Lightly modified
 - Oxidation, Deamidation
- Error Tolerant search to look for additional modifications
- Search using Mascot 2.6
- Search using Mascot 2.7 default permutation parameters
 - MaxPepNumModifiedSites & MaxPepModArrangements
 - As per first pass search
 - MaxPepNumVarMods
 - +1 ET mod

MASCOT : *Modifications*

© 2020 Matrix Science



One of the consequences of these changes are that we can speed up error tolerant searches. To test this, I've taken a subset of a mouse label free dataset available from the PRIDE public repository and produced a peaklist of approximately 20,000 MS/MS spectra. The samples are lightly modified – the information in PRIDE specified Oxidation and Deamidation as variable modification, and Carbamidomethylation as a fixed modification. Then I've searched the peaklists using Mascot 2.6 and Mascot 2.7 using the default permutation parameters.

When Mascot 2.7 carries out an error tolerant search, the MaxPepNumModifiedSites and MaxPepModArrangements settings are applied directly to the error tolerant pass. The MaxPepNumVarMods setting applies only to the first pass search, so we allow for 1 additional error tolerant modification to be applied in the second pass search.

Example 1: Comparison between Mascot 2.6 and 2.7

| Mascot Version | Total Matches | ET matches | Total search time (min) |
|----------------|---------------|------------|-------------------------|
| 2.6 | 4437 | 629 | 95 |
| 2.7 | 4442 | 635 | 79 |

17% faster

First pass search takes ~ 2 minutes

MASCOT : *Modifications*

© 2020 Matrix Science

MATRIX
SCIENCE

The search had a total of 4437 PSMs above the homology threshold, including 629 error tolerant matches, and took 95 minutes with Mascot 2.6

With Mascot 2.7, you can see we have a very slight increase in the number of significant and error tolerant matches, but the results are essentially the same. However, the search was significantly faster, taking 79 minutes – that’s 17% faster.

One of the main reasons for this is the limit on the number of modified sites applied by the MaxPepNumModifiedSites setting. In Mascot 2.6, there was no limit, so, especially on longer peptides, the error tolerant search pass can end up testing a lot of combinations of modifications which don’t actually yield reliable results and are therefore never reported. The MaxPepNumModifiedSites setting of 5 in Mascot 2.7 is preventing this, so the search takes less time.

Example 1: Speeding up Mascot 2.7

- Search 1: MaxPepModArrangements 32
- Search 2: MaxPepNumModifiedSites 3
- Search 3: MaxPepNumVarMods 2
- Search 4: MaxPepNumVarMods 2
MaxPepNumModifiedSites 3
MaxPepModArrangements 32

This particular sample is not very heavily modified – so I should be able to get the same (or very similar) error tolerant search results using narrower permutation settings. Doing this should also reduce the search time. To test this, I've repeated and timed the search on Mascot 2.7 using a range of different permutation settings – where a parameter isn't specified, the default value is used.

| Mascot 2.7 Default | | | | | Mascot 2.7 Search 4 | | | | |
|---------------------------------|-------------|----------|--------|---------------|---------------------------------|-------------|----------|--------|---------------|
| Modification | Delta | Type | Site | Total matches | Modification | Delta | Type | Site | Total matches |
| Carbamidomethyl | 57.021464 | fixed | C | 1054 | Carbamidomethyl | 57.021464 | fixed | C | 1054 |
| Deamidated | 0.984016 | variable | N | 246 | Deamidated | 0.984016 | variable | N | 244 |
| Deamidated | 0.984016 | variable | Q | 112 | Deamidated | 0.984016 | variable | Q | 112 |
| Guandinyl | 42.021792 | ET | N-term | 88 | Guandinyl | 42.021792 | ET | N-term | 88 |
| Non-specific cleavage | | ET | - | 84 | Non-specific cleavage | | ET | - | 84 |
| Gly->Pro | 40.0313 | ET | G | 43 | Gly->Pro | 40.0313 | ET | G | 43 |
| Carbamyl | 43.005816 | ET | N-term | 29 | Carbamyl | 43.005816 | ET | N-term | 29 |
| Glu->pyro-Glu | -17.020632 | ET | N-term | 28 | Glu->pyro-Glu | -17.020632 | ET | N-term | 28 |
| Oxidation | 15.994915 | variable | M | 25 | Oxidation | 15.994915 | variable | M | 25 |
| Xle->Aan | 0.958863 | ET | L | 23 | Xle->Aan | 0.958863 | ET | L | 23 |
| Xle->Aan | 0.958863 | ET | L | 18 | Xle->Aan | 0.958863 | ET | L | 18 |
| CarbamidomethylDTT | 209.018035 | ET | C | 16 | CarbamidomethylDTT | 209.018035 | ET | C | 16 |
| Dehydrated | -18.010565 | ET | T | 15 | Dehydrated | -18.010565 | ET | T | 15 |
| Trimethyl | 42.046954 | ET | N-term | 14 | Trimethyl | 42.046954 | ET | N-term | 14 |
| Val->Ile | -2.01565 | ET | V | 11 | Val->Ile | -2.01565 | ET | V | 11 |
| Arg-loss | -156.101098 | ET | C-term | 11 | Arg-loss | -156.101098 | ET | C-term | 11 |
| Ammonia-loss | -17.020632 | ET | N-term | 11 | Ammonia-loss | -17.020632 | ET | N-term | 11 |
| Met-loss-Acetyl | -89.029905 | ET | N-term | 10 | Met-loss-Acetyl | -89.029905 | ET | N-term | 10 |
| Dehydrated | -18.010565 | ET | S | 10 | Dehydrated | -18.010565 | ET | S | 10 |
| Delta-H3C(C)(2)O(1) | 58.941865 | ET | C | 9 | Delta-H3C(C)(2)O(1) | 58.941865 | ET | C | 9 |
| Glu->Gln | -0.984016 | ET | E | 7 | Glu->Gln | -0.984016 | ET | E | 7 |
| Phe->Val | -18.025821 | ET | F | 7 | Phe->Val | -18.025821 | ET | F | 7 |
| Acetyl | 42.010565 | ET | S | 7 | Acetyl | 42.010565 | ET | S | 7 |
| Didehydro | -2.01565 | ET | T | 6 | Didehydro | -2.01565 | ET | T | 6 |
| Thr->Cys | 1.961506 | ET | T | 6 | Thr->Cys | 1.961506 | ET | T | 6 |
| Acetyl | 42.010565 | ET | N-term | 6 | Acetyl | 42.010565 | ET | N-term | 6 |
| Asp->Asn | -0.984016 | ET | D | 5 | Didehydro | -2.01565 | ET | T | 5 |
| Dehydrated | -18.010565 | ET | D | 5 | Asp->Asn | -0.984016 | ET | D | 5 |
| Glu->pyro-Glu+Methyl2H(2)13C(1) | -0.979001 | ET | N-term | 5 | Dehydrated | -18.010565 | ET | D | 5 |
| Glu->pyro-Glu | -18.010565 | ET | N-term | 5 | Glu->pyro-Glu+Methyl2H(2)13C(1) | -0.979001 | ET | N-term | 5 |
| Thr->Val | -1.979265 | ET | T | 4 | Glu->pyro-Glu | -18.010565 | ET | N-term | 5 |
| Propyl | 42.04695 | ET | E | 4 | Thr->Val | -1.979265 | ET | T | 4 |
| Carbamyl | 13.979265 | ET | A | 4 | Propyl | 42.04695 | ET | E | 4 |
| Guandinyl | 42.021798 | ET | K | 4 | Carbamyl | 13.979265 | ET | A | 4 |
| Dehydrated | -18.010565 | ET | Y | 4 | Guandinyl | 42.021798 | ET | K | 4 |
| Ethanedithiol | 75.980527 | ET | S | 4 | Dehydrated | -18.010565 | ET | Y | 4 |
| Glu->Met | 1.997892 | ET | E | 4 | Ethanedithiol | 75.980527 | ET | S | 4 |
| Phospho | 79.960331 | ET | S | 4 | Glu->Met | 1.997892 | ET | E | 4 |
| ISO_2+2_ion | -15.010877 | ET | N-term | 3 | Phospho | 79.960331 | ET | S | 4 |
| HexNAc(1)Hex(1) | 349.137281 | ET | S | 3 | ISO_2+2_ion | -15.010877 | ET | N-term | 3 |
| HexNAc(1)Hex(1) | 349.137281 | ET | T | 3 | HexNAc(1)Hex(1) | 349.137281 | ET | S | 3 |
| SulfurDioxide | 63.9619 | ET | C | 3 | HexNAc(1)Hex(1) | 349.137281 | ET | T | 3 |
| Didehydro | -2.01565 | ET | S | 3 | SulfurDioxide | 63.9619 | ET | C | 3 |
| Didehydro | -2.01565 | ET | Y | 3 | Didehydro | -2.01565 | ET | S | 3 |
| Amidated | -0.984017 | ET | C-term | 3 | Didehydro | -2.01565 | ET | Y | 3 |
| Xink-DTBP[87] | 87.01427 | ET | K | 3 | Amidated | -0.984017 | ET | C-term | 3 |
| Ethanolamine | 43.042199 | ET | E | 3 | Xink-DTBP[87] | 87.01427 | ET | K | 3 |
| Carbamyl | 43.005814 | ET | Y | 3 | Ethanolamine | 43.042199 | ET | E | 3 |
| Glu->Aan | -14.999666 | ET | E | 2 | Carbamyl | 43.005814 | ET | Y | 3 |
| Carbamoylmethyl | 58.005479 | ET | C | 2 | Glu->Aan | -14.999666 | ET | E | 2 |
| Ala->Aan | 43.005814 | ET | A | 2 | Carbamoylmethyl | 58.005479 | ET | C | 2 |
| Oxidation | 15.994915 | ET | P | 2 | Ala->Aan | 43.005814 | ET | A | 2 |
| Thr->Ala | -30.010565 | ET | T | 2 | Oxidation | 15.994915 | ET | P | 2 |
| azole | -30.030215 | ET | S | 2 | Thr->Ala | -30.010565 | ET | T | 2 |
| Deoxy | -15.994915 | ET | S | 2 | azole | -30.030215 | ET | S | 2 |
| Sulfo | 79.956815 | ET | S | 2 | Deoxy | -15.994915 | ET | S | 2 |
| Xle->Asp | 1.942879 | ET | I | 2 | Sulfo | 79.956815 | ET | S | 2 |
| | | | | | Xle->Asp | 1.942879 | ET | I | 2 |

4442 PSMs

4443 PSMs

MASCOT : Modifications

© 2020 Matrix Science



Once the searches have completed, you can see the modifications which have been identified, and the number of PSMs containing the modification by expanding the ‘Modification statistics for all protein families’ section on the protein family report.

As you can see, the list of the top most identified error tolerant variable modifications identified by the default and search 4 permutation settings are almost identical – we actually get one extra match from the lower permutation settings. This is probably due to a marginal match becoming significant due to the reduced search space.

Example 1: Results

| Search | ModArrangements | NumModifiedSites | NumVarMods | Speed improvement (%) |
|--------|-----------------|------------------|------------|-----------------------|
| 1 | 32 | --- | --- | 14 |
| 2 | --- | 3 | --- | 21 |
| 3 | --- | --- | 2 | 0 |
| 4 | 32 | 3 | 2 | 28 |

MASCOT : *Modifications*

© 2020 Matrix Science



We get very similar results from all the searches on both the standard, first pass, search and the error tolerant pass. Reducing the maximum number of modification arrangements tested to 32, gives us a 14% improvement in the search speed, while reducing the number of modified sites allowed to 3 gives us a 21% improvement. In both cases, that's because we're reducing the search space, which has quite a big impact on the time taken for the error tolerant search.

Reducing the maximum number of different variable modifications per peptide doesn't improve the search speed. In this case, that is expected because we've only got 2 variable modifications selected in the search settings, so changing this won't have any effect on the search space.

Lowering all three settings gives us the fastest search, which is now 28% faster than using the default settings – this is mainly due to the combination of the reduced number of mod arrangements and modified sites.

So, when you're search datasets which are not highly modified with Mascot 2.7, you may wish to consider lowering the permutation settings to speed up searching – especially if you're doing an error tolerant pass.

Example 2: Middle down dataset

- **Human Histone H4 dataset (PXD008296)**
 - Chosen a single raw file
- **Digested with Asp-N**
- **Selected first peptide (23 residues)**
 - SGRGKGGKGLGKGGAKRHRKVLK
 - Highly modified
- **Looking for different modification patterns**

MASCOT : *Modifications*

© 2020 Matrix Science

 **MATRIX
SCIENCE**

Our second example combines the two reasons you might want to increase the variable modification permutation limits.

I took a single raw file from a human histone H4 middle down dataset available on the PRIDE repository. The authors had digested the sample with Asp-N and selected the first 23 residue peptide. Being a Histone protein, this peptide is highly modified, and the authors were looking for changes in modification patterns across the cell cycle.

Example 2: Middle down dataset

- **Variable Modifications**
 - Acetyl (K), Acetyl (Protein N-term), Phospho (ST), Dimethyl (K), Methyl (K), Methyl (R), Trimethyl (K)
- **SGRGKGGKGLGKGGAKRHRKVL**
 - 1 x S, 4 x R, 5 x K
- **Default permutation settings not sufficient**

The variable modifications for the search were taken from the paper – as you can see the peptide contains a large number of possible modification sites, particularly for acetylation and methylation. Because of this, it's clear that the default permutation settings won't be sufficient to give us accurate modification analysis.

Example 2: Middle down dataset

- Search with Mascot 2.6
- Search using Mascot 2.7 with the following permutation settings:
 - MaxPepNumModifiedSites: 11
 - MaxPepNumVarMods: 7
 - MaxPepModArrangements:
 - 64
 - 128
 - 256
 - 512

MASCOT : *Modifications*

© 2020 Matrix Science



I processed the raw file in Mascot Distiller and exported an MGF file with the fragment ion masses decharged to 1+, then searched using Mascot 2.6 and Mascot 2.7. For the Mascot 2.7 search I changed the permutation settings. I set MaxPepNumModifiedSites to 11 and MaxPepNumVarMods to 7, so that all possible residues could be modified and allowing for all the variable modifications. I then searched using a range of values for MaxPepModArrangements, from 64 up to 512.

Example 2: Counts of proteoforms

Mascot 2.6

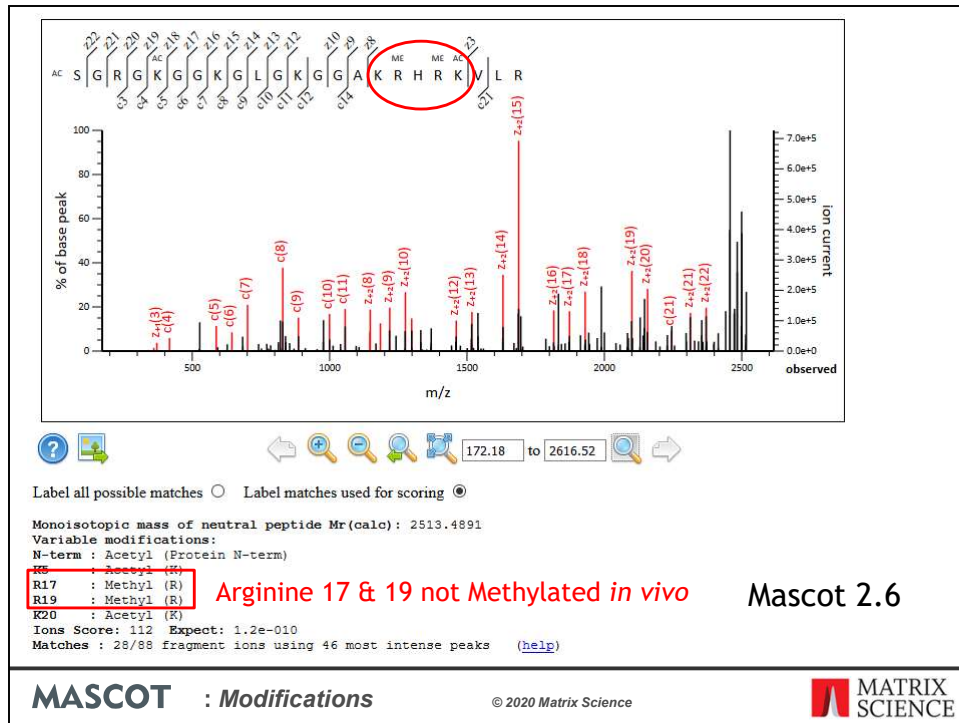
| | Number of proteoforms |
|--|-----------------------|
| | 81 |

Mascot 2.7

| MaxPepModArrangements | Number of proteoforms |
|-----------------------|-----------------------|
| 64 | 106 |
| 128 | 109 |
| 256 | 109 |
| 512 | 109 |

With Mascot 2.6, we identified 81 distinct proteoforms – that is distinct modification patterns.

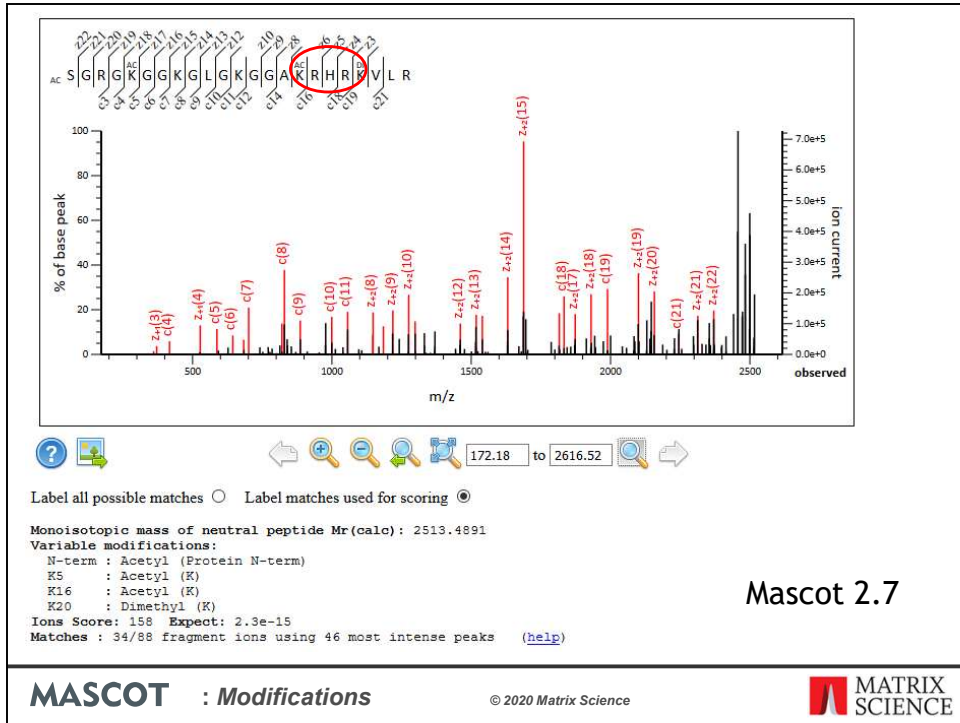
As you can see, using Mascot 2.7 to do a deeper search into the variable modifications, we have identified significantly more possible proteoforms, although there was no need to go above 128 arrangements – we hit a maximum number of proteoforms of 109 and increasing beyond 128 will have done nothing other than slow the search down.



Lets take a look at a few examples of the differences between the results for Mascot 2.6 and 2.7. In all cases, the Mascot 2.7 results are taken from the search where we used 128 modification arrangements.

Here we have a very good match from Mascot 2.6 with an ions score of 112. It looks reliable – however, there is a problem.

Arginines 17 and 19 are not methylated in vivo, and indeed, if we look at that region of the sequence ladder we have no matching fragments in that region. What appears to have happened is that we've got the clustering of modifications that I mentioned was an issue with the sliding window method used in Mascot 2.6 for peptides with a lot of variable modification permutations. I think we'd have to conclude that, while we have the correct peptide sequence here, the modification pattern reported is incorrect.



Mascot 2.7

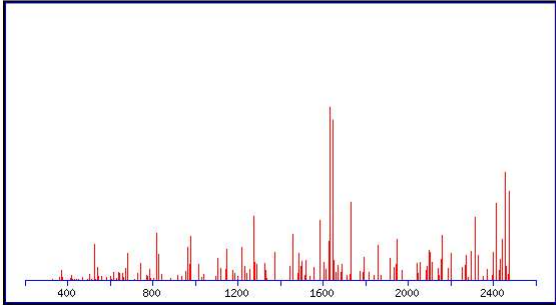
Here we have the best match to the same peaklist in Mascot 2.7 – and we have an even better match. The ions score is now increased to 158, and the modification pattern is much more believable. If we look at the sequence ladder, the region around R17 and R19 is now covered.

MATRIX SCIENCE Mascot Search Results

Peptide View

No match to Query 785: 2472.482124 from(413.087630.6+) intensity(10917784.0000) scans(594)
Title: 442: Scan 594 (rt=50.088) [D:\PEmeryTemp\ASMS_thoughts\8296H4_thy_2_1_10A.raw]
Data file fda3ed686c833002__H4_thy_2_1_10A.raw-1.mgf

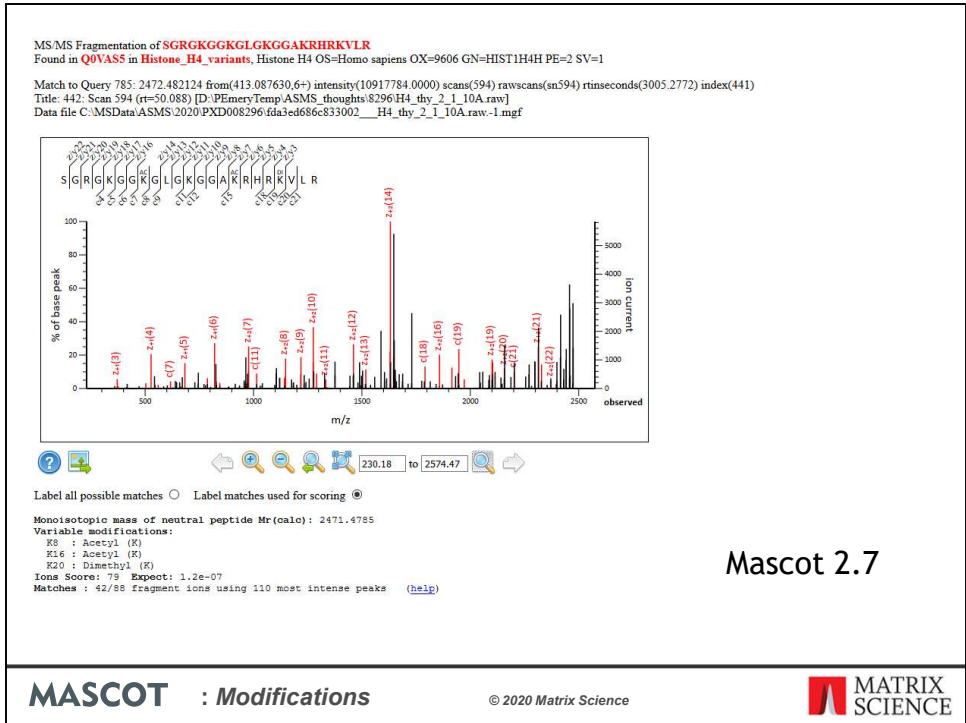
Click mouse within plot area to zoom in by factor of two about that point
Or, Plot from to Da
Show Y-axis



Mascot 2.6

MASCOT : Modifications © 2020 Matrix Science **MATRIX SCIENCE**

Here we have an example where Mascot 2.6 failed to find any match for the peaklist



But with Mascot 2.7 we've got a good match to the same peaklist. Presumably the correct modification pattern was not tested by Mascot 2.6, preventing us from getting a match

Summary

- **Variable Modification permutation has changed in Mascot 2.7**
- **Controlled by three new parameters**
- **Reduce the search depth**
 - Faster error tolerant searches
 - Speed up searches of lightly modified samples
 - Or where site localisation less important
- **Increase the search depth**
 - Where site localisation matters
 - Or for highly modified samples

So, to summarise:

Permutation of variable modification arrangements has changed significantly in Mascot 2.7. This has been done to address some short comings in how earlier versions of Mascot handled this for peptides with large numbers of possible modification arrangements.

The new variable modification iterator is controlled by three parameters which can be set globally and overridden locally. This gives you a large amount of control over the depth of search when Mascot assigns variable modifications.

By controlling these parameters, you can either reduce the depth of the search. This will speed up searches of lightly modified samples without affecting your results. Or you can increase the depth of the search, which is important if you are interested in site localisation or in studying highly modified peptides.