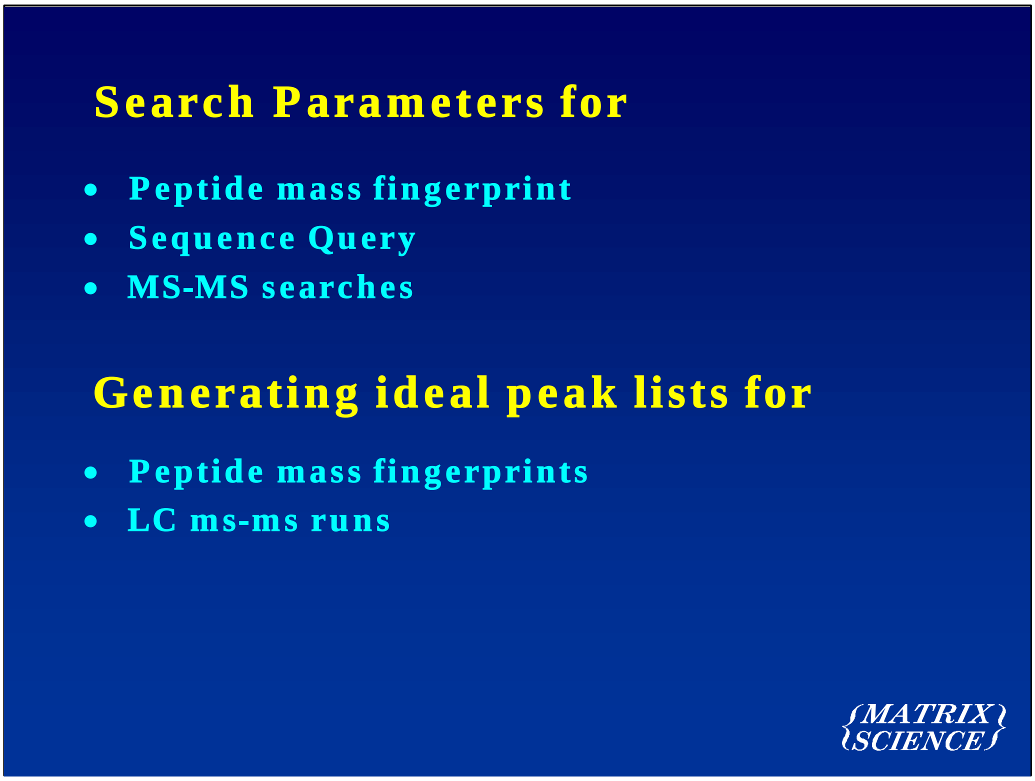
Mascot Search Parameters

MATRIX SCIENCE

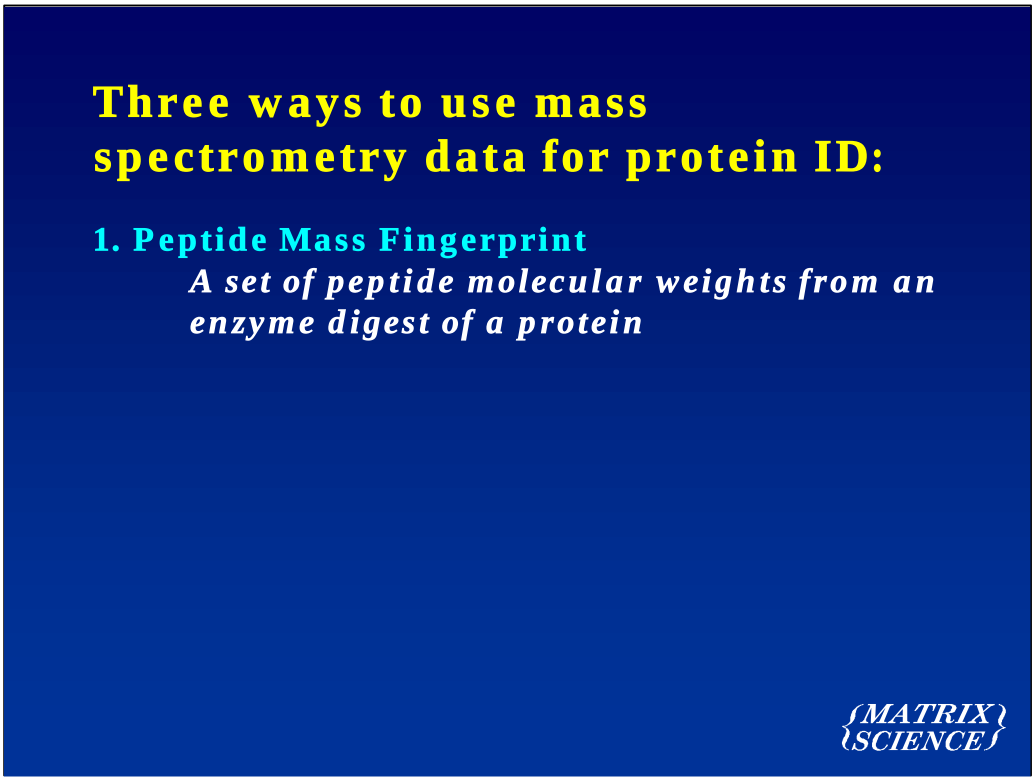


Search Parameters for

- Peptide mass fingerprint
- Sequence Query
- MS-MS searches

Generating ideal peak lists for

- Peptide mass fingerprints
- LC ms-ms runs



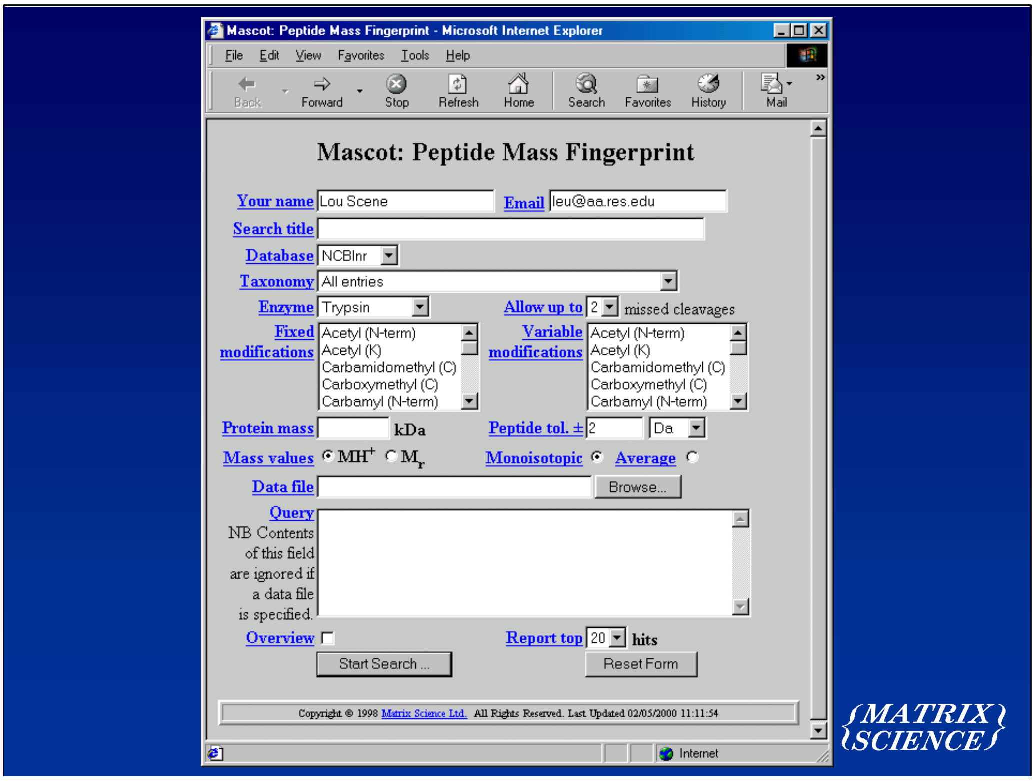
Three ways to use mass spectrometry data for protein ID:

1. Peptide Mass Fingerprint

A set of peptide molecular weights from an enzyme digest of a protein

There are three proven ways of using mass spectrometry data for protein identification.

The first of these is known as a peptide mass fingerprint. This was the first method to be developed, and is in many ways the simplest approach.



Mascot: Peptide Mass Fingerprint - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Forward Stop Refresh Home Search Favorites History Mail
Internet
MATRIX SCIENCE

Mascot: Peptide Mass Fingerprint

Your name: Lou Scene

Email: leu@aa.res.edu

Search title:

Database: NCBInr

Taxonomy: All entries

Enzyme: Trypsin

Allow up to 2 missed cleavages

Fixed modifications: Acetyl (N-term), Acetyl (K), Carbamidomethyl (C), Carboxymethyl (C), Carbamyl (N-term)

Variable modifications: Acetyl (N-term), Acetyl (K), Carbamidomethyl (C), Carboxymethyl (C), Carbamyl (N-term)

Protein mass: kDa

Peptide tol. ± 2 Da

Mass values: ● MH^+ ○ M_r

Monoisotopic ● Average ○

Data file: Browse...

Query: NB Contents of this field are ignored if a data file is specified.

Overview □

Report top 20 hits

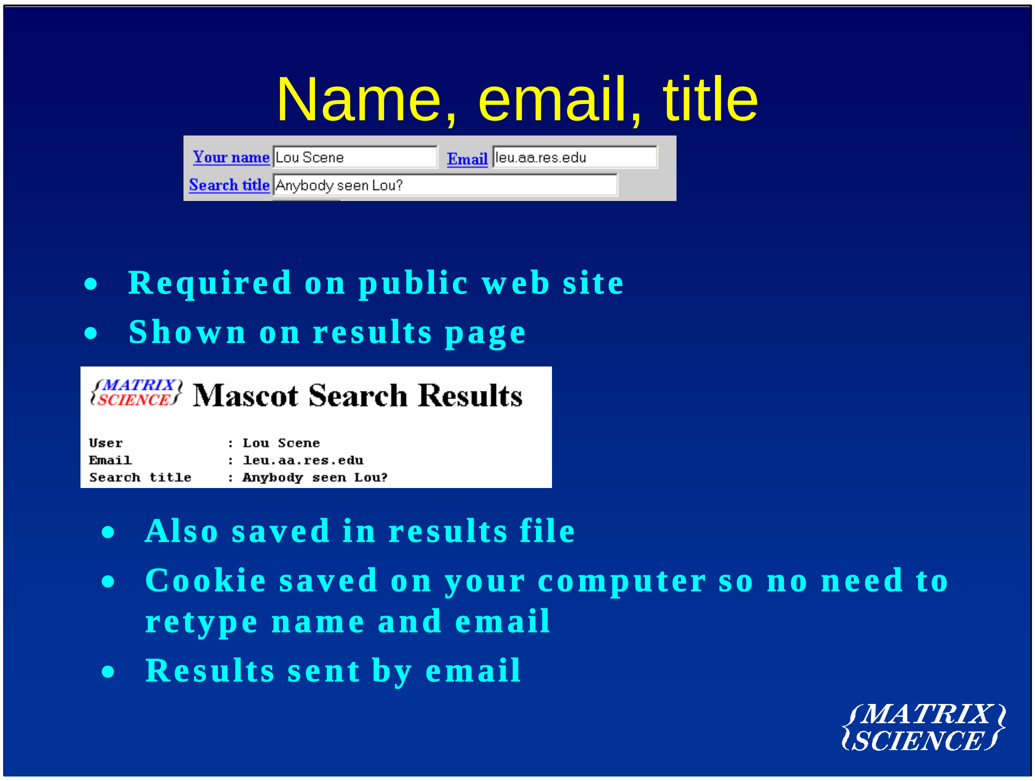
Start Search ... Reset Form

Copyright © 1998 Matrix Science Ltd. All Rights Reserved. Last Updated 02/05/2000 11:11:54

This is the appearance of the search form for a peptide mass fingerprint.

The parameters editor tab in Mascot Daemon looks very similar.

There is on line help for all of the parameters - simply click on the blue hyperlink for extensive help on each item.



Name, email, title

Your name: Lou Scene Email: leu.aa.res.edu
Search title: Anybody seen Lou?

- Required on public web site
- Shown on results page

{MATRIX SCIENCE} Mascot Search Results

```
User           : Lou Scene
Email          : leu.aa.res.edu
Search title   : Anybody seen Lou?
```

- Also saved in results file
- Cookie saved on your computer so no need to retype name and email
- Results sent by email

{MATRIX SCIENCE}

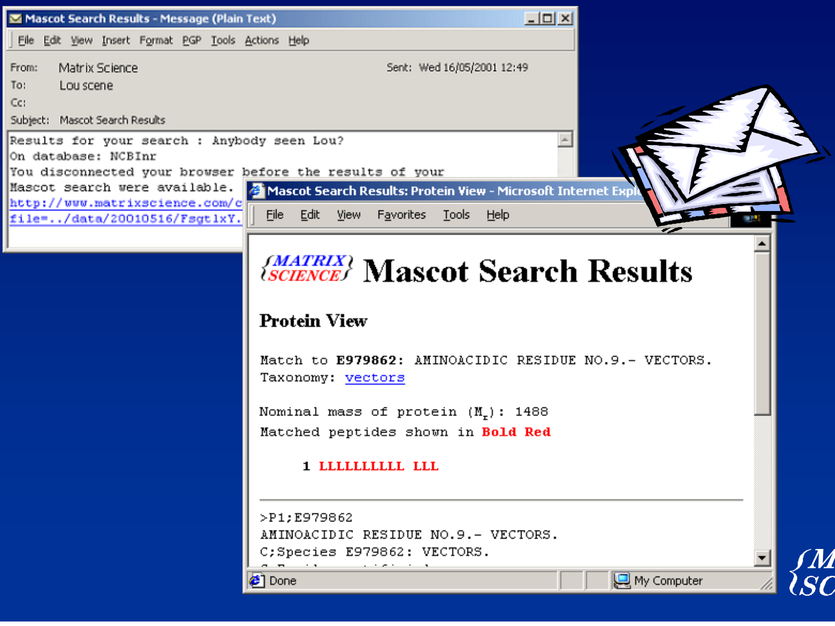
On our public web site, you must enter your name and email address. On the intranet version, these fields can be left blank, but we recommend that you continue to supply this information because it can be used to track results.

The information that you enter will appear on the results report and it is also saved in the results files

Your name and email address are saved in a cookie on your PC, so you don’t have to type them in each time.

On the web site, the reason for requesting an email address is so that results can be emailed to you if the connection between your browser and our server gets broken during a long search. If you are unable to use your company or institution email address, you can always set up a temporary email address.

Results by email


Mascot Search Results - Message (Plain Text)
File Edit View Insert Format PGP Tools Actions Help

From: Matrix Science Sent: Wed 16/05/2001 12:49
To: Lou scene
Cc:
Subject: Mascot Search Results

```
Results for your search : Anybody seen Lou?
On database: NCBInr
You disconnected your browser before the results of your
Mascot search were available.
http://www.matrixscience.com/c
file=../data/20010516/FsgtlxY.
```

Mascot Search Results: Protein View - Microsoft Internet Expl
File Edit View Favorites Tools Help
Done
My Computer

Mascot Search Results

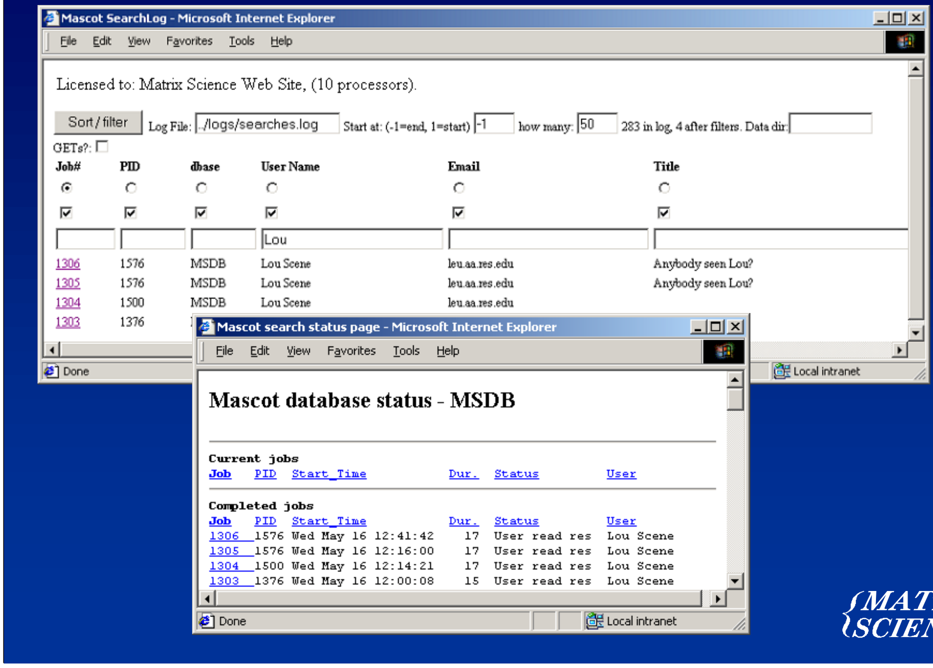
Protein View

```
Match to E979862: AMINOACIDIC RESIDUE NO.9.- VECTORS.
Taxonomy: vectors

Nominal mass of protein (Mr): 1488
Matched peptides shown in Bold Red

     1 LLLLLLLLLL LLL

>P1;E979862
AMINOACIDIC RESIDUE NO.9.- VECTORS.
C;Species E979862: VECTORS.
```

Receiving results by email for a peptide mass fingerprint search should be rare, because the searches are so fast. However, with a long MS-MS search, it is possible for the connection to get broken. If this happens, then, assuming that you have entered a valid email address, you should receive an email soon after the search has finished. The email will have the search title, and a link to the results on our public server, or your intranet server.

You can then open up the results, and click on all the links as you would normally.

Search Log and status screens


Mascot SearchLog - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Sort / filter
Done
Local intranet
Mascot search status page - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Done
Local intranet

Licensed to: Matrix Science Web Site, (10 processors).

Log File: ../logs/searches.log Start at: (-1=end, 1=start) -1 how many: 50 283 in log, 4 after filters. Data dir:

GETs?: □

Job#	PID	dbase	User Name	Email	Title
●	○	○	○	○	○
☑	☑	☑	☑	☑	☑
			Lou		
1306	1576	MSDB	Lou Scene	leu.aa.res.edu	Anybody seen Lou?
1305	1576	MSDB	Lou Scene	leu.aa.res.edu	Anybody seen Lou?
1304	1500	MSDB	Lou Scene	leu.aa.res.edu	
1303	1376				

Mascot database status - MSDB

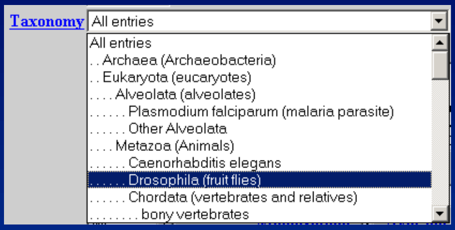
```
Current jobs
Job    PID  Start_Time                Dur.  Status          User

Completed jobs
Job    PID  Start_Time                Dur.  Status          User
1306   1576 Wed May 16 12:41:42         17  User read res   Lou Scene
1305   1576 Wed May 16 12:16:00         17  User read res   Lou Scene
1304   1500 Wed May 16 12:14:21         17  User read res   Lou Scene
1303   1376 Wed May 16 12:00:08         15  User read res   Lou Scene
```

These administration screens are hidden on the public web site.

If you have an in-house copy of Mascot, your name will appear in the status screens. This can be useful if many different people are using the system, and you want to track a search.

The search log contains all the searches that have been performed. You can filter this list to show all the searches that you have performed - or look for search with a particular title.

Taxonomy


Taxonomy
All entries
All entries
. . Archaea (Archaeobacteria)
. . Eukaryota (eucaryotes)
. . . . Alveolata (alveolates)
. . . . . . Plasmodium falciparum (malaria parasite)
. . . . . . Other Alveolata
. . . . Metazoa (Animals)
. . . . . . Caenorhabditis elegans
. . . . . . Drosophila (fruit flies)
. . . . . . Chordata (vertebrates and relatives)
. . . . bony vertebrates

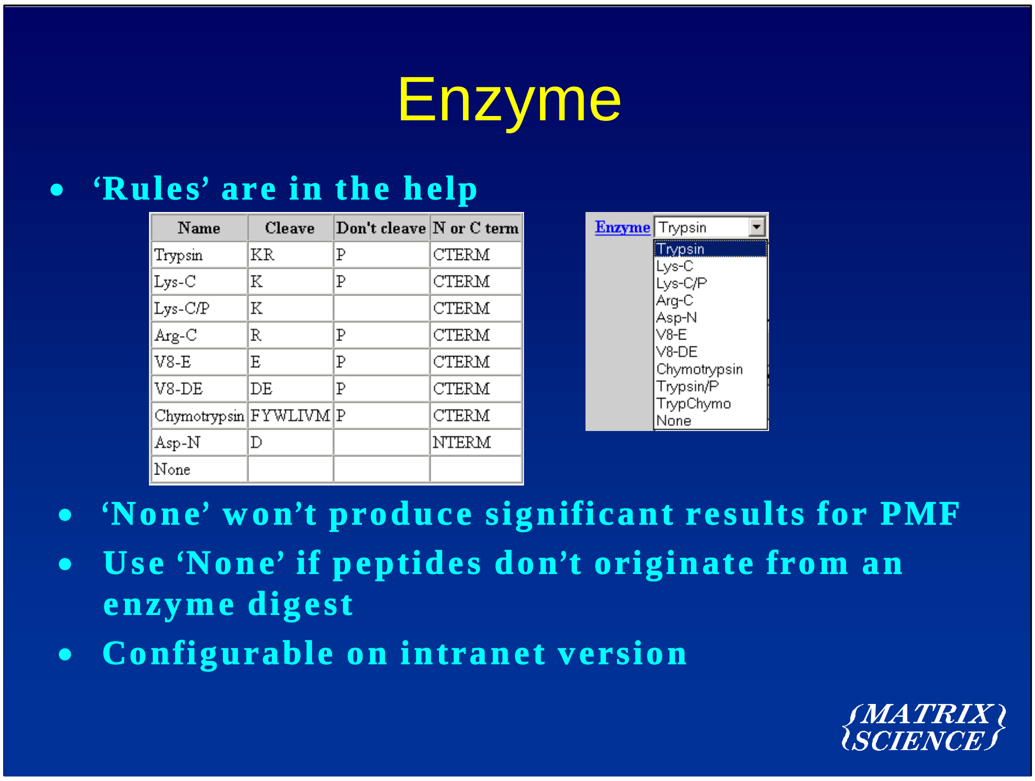
- **Useful for speeding up searches**
- **Useful for limiting number of matches in result**
- **May be no protein match for selected species**
- **Not totally reliable, but getting better with NCBI**
- **Drop down list configurable**

The next item on the form is the taxonomy filter. This is useful for speeding up searches and for limiting the number of matches in a result page. However, you should be aware of two limitations:

1 - There may not yet be a database entry for your particular protein in the species that you have selected - but somebody may have submitted an identical or nearly identical protein for a similar species.

2 - The parsing of taxonomy species from the text based files is only about 99% reliable.

On an in-house copy of Mascot, the list in the drop down box is configurable.



Enzyme

- ‘Rules’ are in the help

Name	Cleave	Don't cleave	N or C term
Trypsin	KR	P	CTERM
Lys-C	K	P	CTERM
Lys-C/P	K		CTERM
Arg-C	R	P	CTERM
V8-E	E	P	CTERM
V8-DE	DE	P	CTERM
Chymotrypsin	FYWLIVM	P	CTERM
Asp-N	D		NTERM
None			

- ‘None’ won’t produce significant results for PMF
- Use ‘None’ if peptides don’t originate from an enzyme digest
- Configurable on intranet version

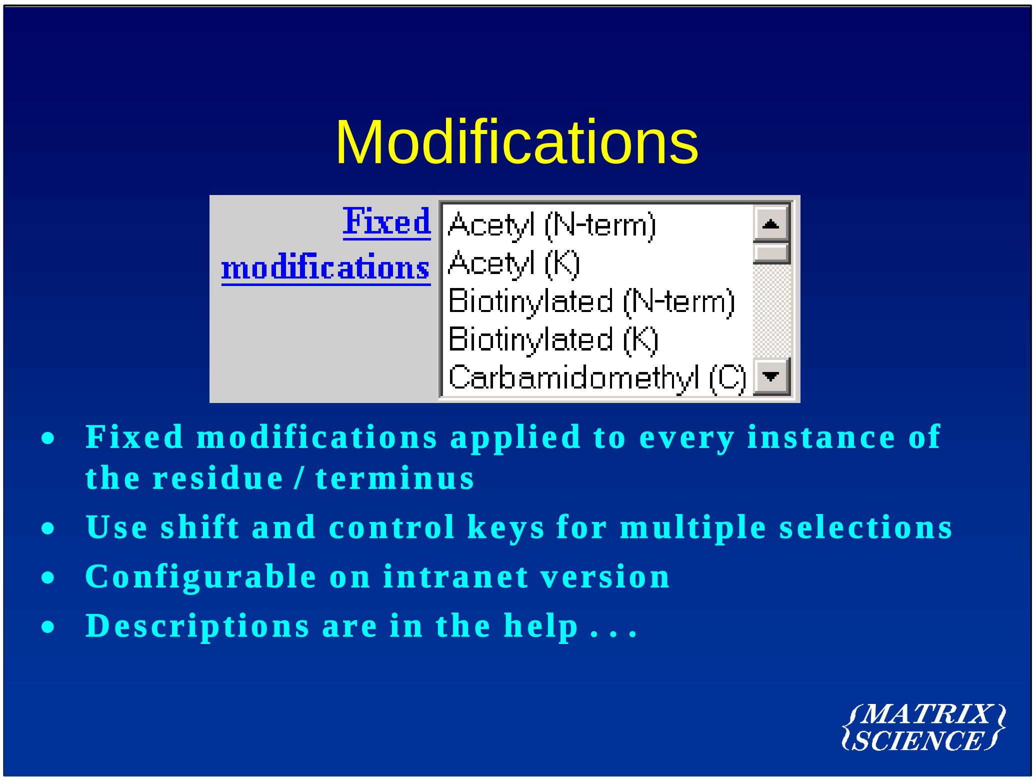
{MATRIX}
{SCIENCE}

You need to select the enzyme which you used to digest your protein.

The cleavage rules for each of the enzymes can be found by clicking on the ‘Enzyme’ link.

You can’t select ‘None’ for a peptide mass fingerprint because it would never result in a significant match. For MS-MS searching and sequence queries, ‘None’ is available, and should be used if you are searching peptides that don’t originate from an enzyme digest.

With an in-house version of Mascot, you can add your own Enzymes to the list by editing a simple text file. If you would like to see additional enzymes on our public web site, please send a request by email.



Modifications

Fixed modifications

- Acetyl (N-term)
- Acetyl (K)
- Biotinylated (N-term)
- Biotinylated (K)
- Carbamidomethyl (C)

- Fixed modifications applied to every instance of the residue / terminus
- Use shift and control keys for multiple selections
- Configurable on intranet version
- Descriptions are in the help . . .

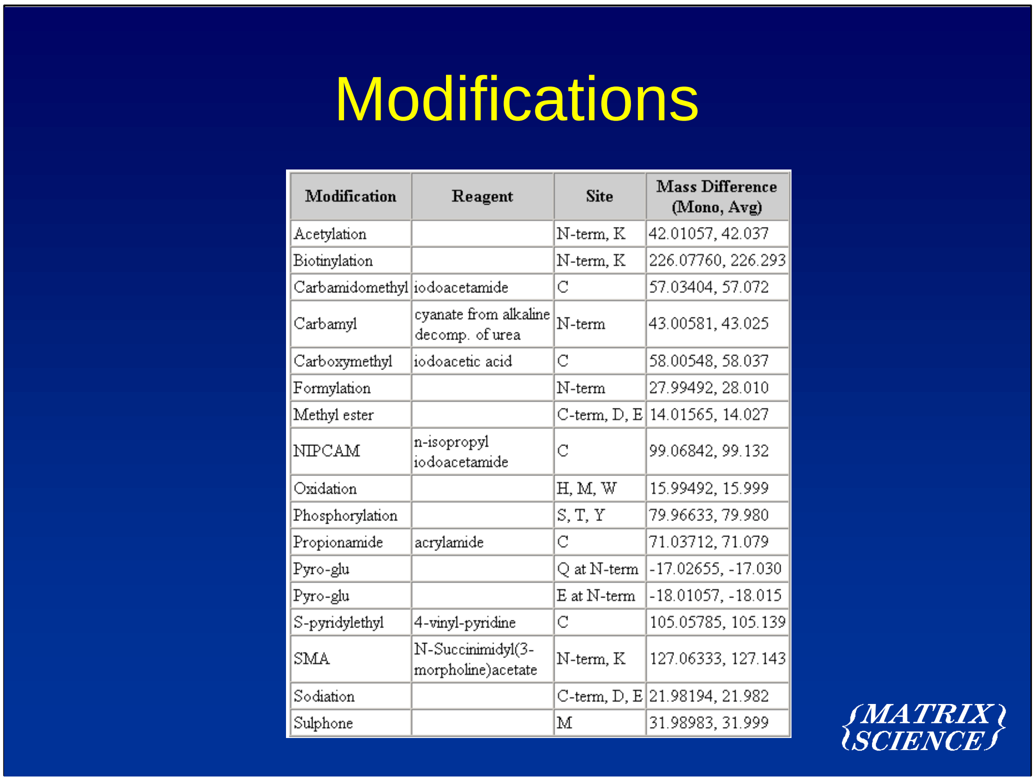
With chemical modifications, such as carbamidomethyl cysteine, residues will be quantitatively modified. This should be selected as a fixed modification.

A fixed modification assumes that every instance of that residue has been modified, so there is no computational overhead to the search, and the score will not be adversely affected in the same way as with variable modifications.

The one question that we probably get asked most often, is “How do I select more than one modification”. It varies a little between browsers, but hold down the control key or shift key to select when clicking on the second and subsequent modifications.

New modifications can be added with an intranet version. We are often prepared to add new modifications to the list on our public web server if you supply us with all the correct information.

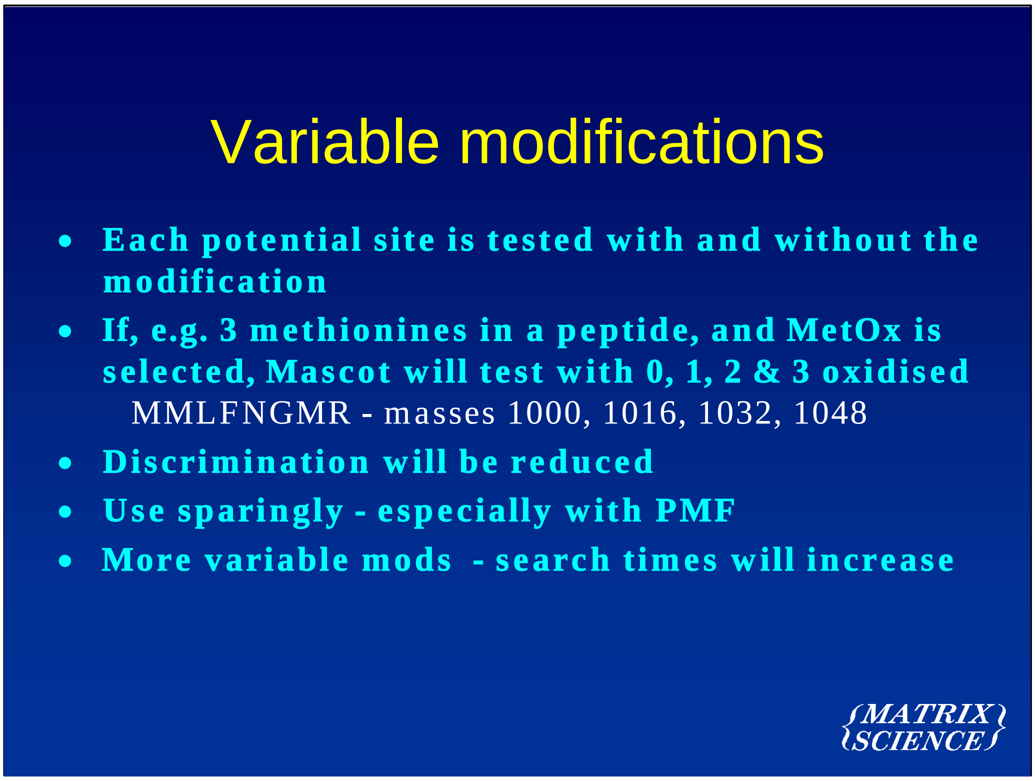
Clicking on the ‘Fixed modifications link’ displays the list of modifications.



Modifications

Modification	Reagent	Site	Mass Difference (Mono, Avg)
Acetylation		N-term, K	42.01057, 42.037
Biotinylation		N-term, K	226.07760, 226.293
Carbamidomethyl	iodoacetamide	C	57.03404, 57.072
Carbamyl	cyanate from alkaline decomp. of urea	N-term	43.00581, 43.025
Carboxymethyl	iodoacetic acid	C	58.00548, 58.037
Formylation		N-term	27.99492, 28.010
Methyl ester		C-term, D, E	14.01565, 14.027
NIPCAM	n-isopropyl iodoacetamide	C	99.06842, 99.132
Oxidation		H, M, W	15.99492, 15.999
Phosphorylation		S, T, Y	79.96633, 79.980
Propionamide	acrylamide	C	71.03712, 71.079
Pyro-glu		Q at N-term	-17.02655, -17.030
Pyro-glu		E at N-term	-18.01057, -18.015
S-pyridylethyl	4-vinyl-pyridine	C	105.05785, 105.139
SMA	N-Succinimidyl(3-morpholine)acetate	N-term, K	127.06333, 127.143
Sodiation		C-term, D, E	21.98194, 21.982
Sulphone		M	31.98983, 31.999

It is worth noting here that we use the name of the modification in the list - not the name of the reagent, so for example, if your cysteines are modified by acrylamide from a gel, you need to enter the modification as propionamide.



Variable modifications

- Each potential site is tested with and without the modification
- If, e.g. 3 methionines in a peptide, and MetOx is selected, Mascot will test with 0, 1, 2 & 3 oxidised
  MMLFNGMR - masses 1000, 1016, 1032, 1048
- Discrimination will be reduced
- Use sparingly - especially with PMF
- More variable mods - search times will increase

MATRIX
SCIENCE

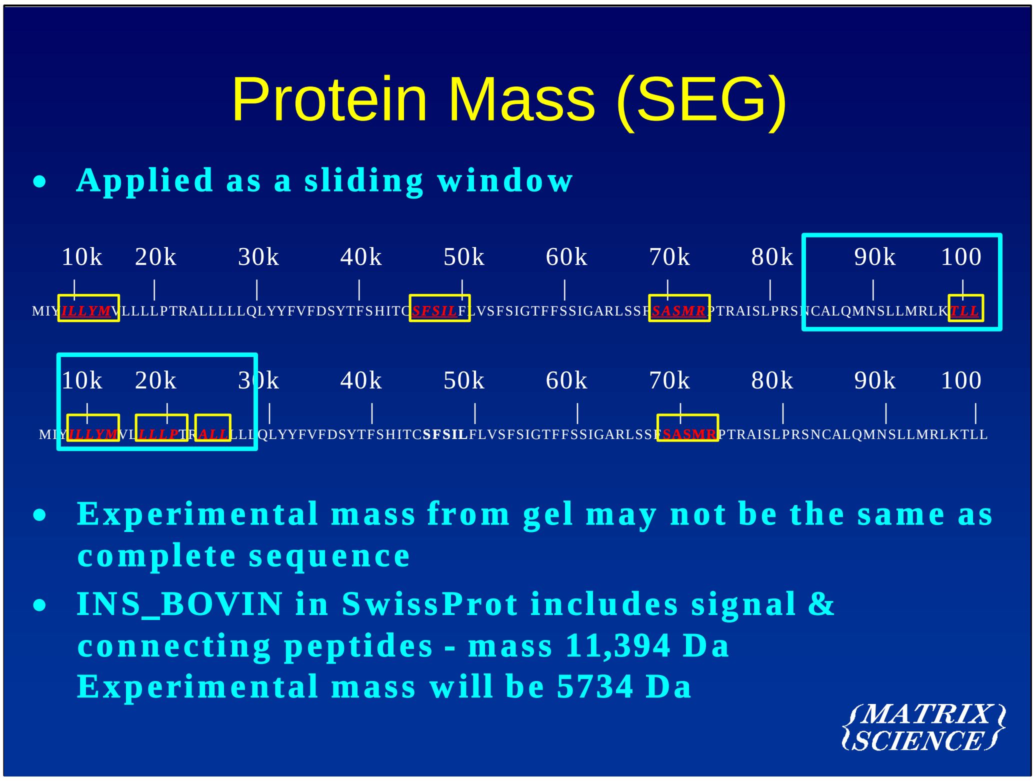
It is important to have a good understanding of what Mascot is doing with variable modifications. Each potential site is tested with and without the modification.

This means, for example, that if there are three methionines in a peptide, and MetOx is selected, then Mascot will try and get a match with 0 oxidised, 1 oxidised, 2 oxidised and 3 oxidised.

This obviously increases the chance of a random match, so if you add an unnecessary variable modification, the score will go down..

So, use modifications sparingly with a peptide mass fingerprint. If you wish to try several modifications, it may be best to try one at a time.

The other thing to note, is that search times will increase, particularly with MS-MS searches.



Protein Mass (SEG)

- **Applied as a sliding window**

10k 20k 30k 40k 50k 60k 70k 80k 90k 100
MIYILLYMVLLLLPTRALLLLLQLYYFVFDSYTFSHITCSFSILFLVSFSIGTFFSSIGARLSSFSASMRPTRAISLPRSNCALQMNSLLMRLKTLL
10k 20k 30k 40k 50k 60k 70k 80k 90k 100
MIYILLYMVLLLLPTRALLLLLQLYYFVFDSYTFSHITCSFSILFLVSFSIGTFFSSIGARLSSFSASMRPTRAISLPRSNCALQMNSLLMRLKTLL

- **Experimental mass from gel may not be the same as complete sequence**
- **INS_BOVIN in SwissProt includes signal & connecting peptides - mass 11,394 Da Experimental mass will be 5734 Da**

{MATRIX}
{SCIENCE}

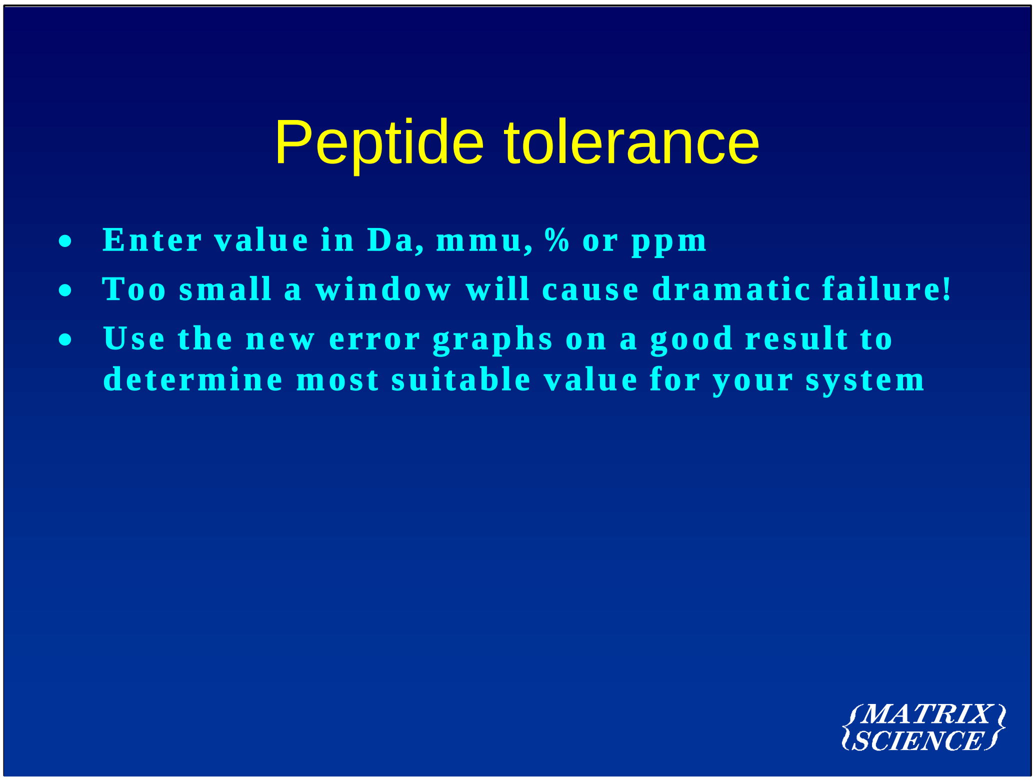
The next item on the form is one of the most misunderstood parameters. The protein mass is not the mass of the complete database entry. We apply protein mass as a sliding window.

For example, assume a protein in the database of 100k Daltons, and four peptides that matched, (shown in red here). If a 20k protein segment mass was applied, then the best match for this protein would only be a single peptide.

If however three of the peptides matched as shown here, and a protein mass of 20k is applied, then 3 peptides would match. The scoring algorithms in Mascot take into account the segment mass, and it is possible in this case that 3 matches from a 20k segment will give a better score than 4 from the complete 100kD protein.

The reason for this approach is that the experimental protein may be shorter than the entry in the database. For example,the entry in Swiss-prot for bovine insulin includes signal and connecting peptides, bringing the mass up to 11,394. An experimentally determined mass would only be 5734, so if we implemented the algorithm using the complete protein molecular weight, you would never get a match.

Also, please note that the mass is in kDa, and not in Daltons.



Peptide tolerance

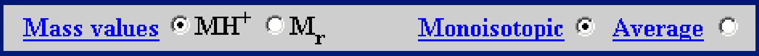
- Enter value in Da, mmu, % or ppm
- Too small a window will cause dramatic failure!
- Use the new error graphs on a good result to determine most suitable value for your system

MATRIX SCIENCE

The peptide tolerance can be entered in Daltons, millimass units (that's units of .001 of a Dalton), percentage, or parts per million.

One of the most common reasons for Mascot failing to get a match is that the peptide tolerance is too tight. In version 1.7, we now have error graphs on the results pages that should help you enter a better tuned value here.

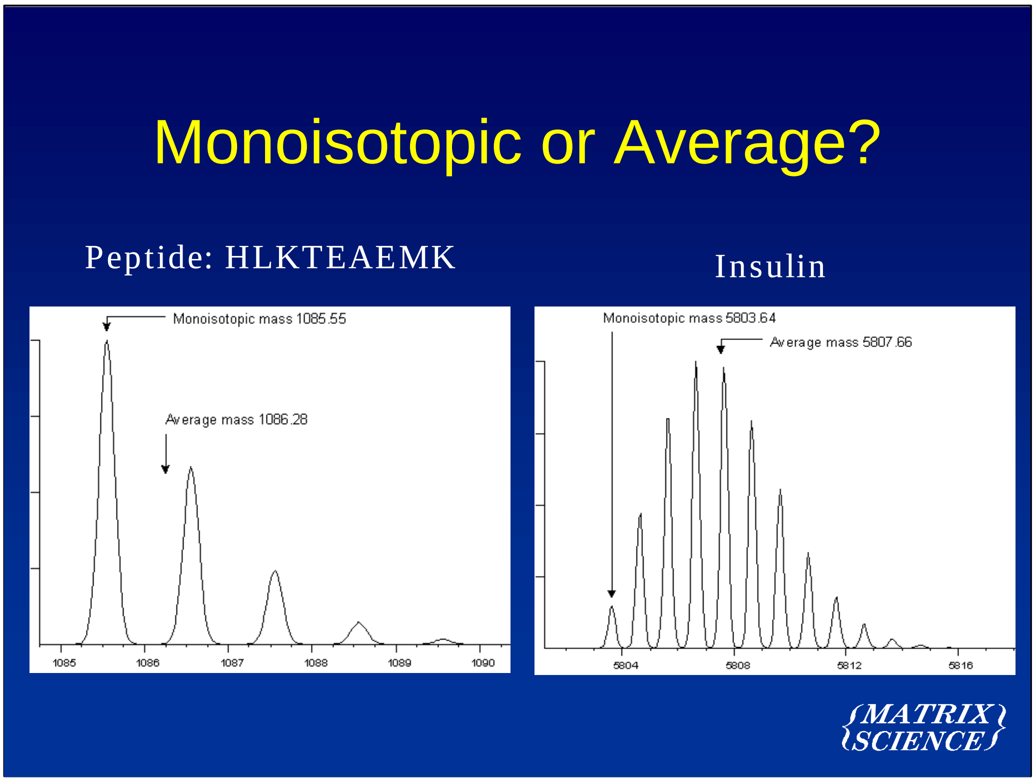
Mass values - MH^+ or M_r



Mass values ● MH^+ ○ M_r Monoisotopic ● Average ○

- Select MH+ if masses include the mass of the charge carrying proton
- Sequence query has default of M_r

{MATRIX}
{SCIENCE}



Monoisotopic or Average?

Peptide: HLKTEAEMK

Insulin

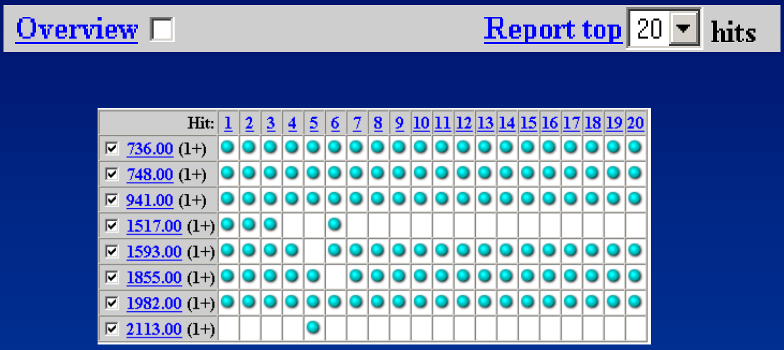
The presence of isotopes at their natural abundances makes it essential to define whether an experimental mass value is an "average" value, equivalent to taking the centroid of the complete isotopic envelope, or a "monoisotopic" value, the mass of the first peak of the isotope distribution.

For peptides and proteins, the difference between an average and a monoisotopic weight is approximately 0.06%. This is a significant difference when even the most modest instruments are capable of measuring the mass of a small peptide with high accuracy. For example, a peptide with an average molecular weight of 1086.28 has a monoisotopic weight of 1085.55. At a mass resolution of 5000, the isotopic envelope looks like this.

To measure a monoisotopic molecular weight requires sufficient mass resolution to resolve the isotopic distribution and sufficient signal to noise to be able to identify the first peak of the envelope with confidence.

Note that the monoisotopic peak is not always the largest peak. For example, insulin at resolution 5000 has this isotopic envelope.

Overview and Report top hits

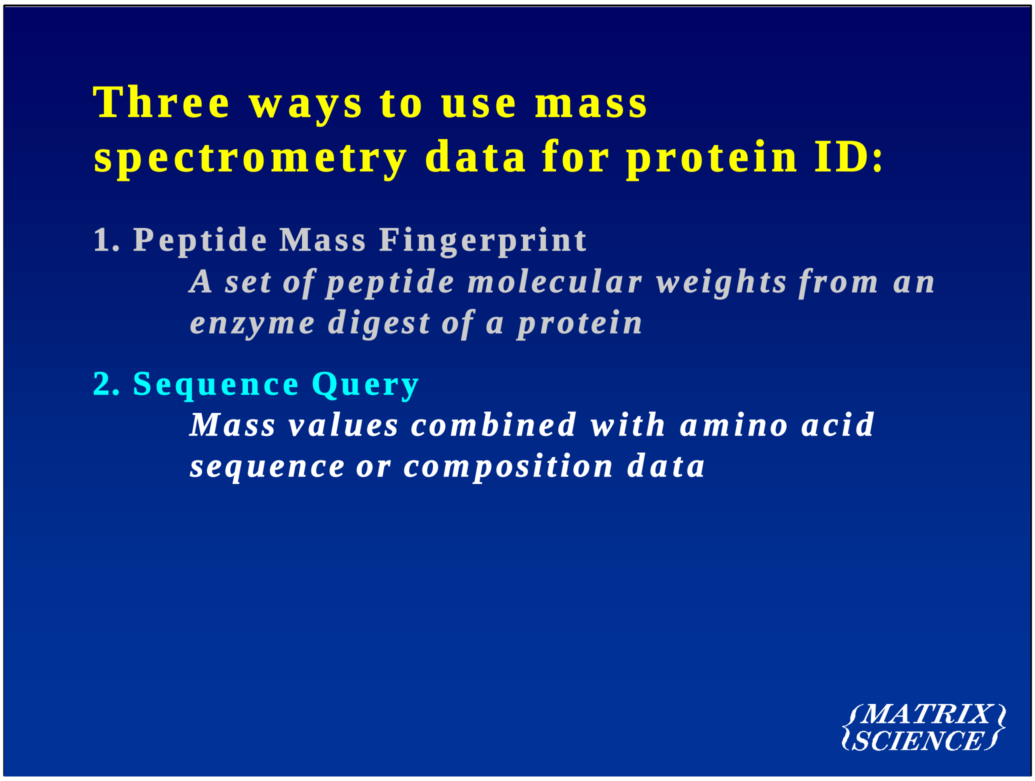


Overview □ Report top 20 ▾ hits

Hit:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
☑ 736.00 (1+)	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
☑ 748.00 (1+)	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
☑ 941.00 (1+)	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
☑ 1517.00 (1+)	●	●	●			●														
☑ 1593.00 (1+)	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
☑ 1855.00 (1+)	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●
☑ 1982.00 (1+)	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
☑ 2113.00 (1+)					●															

{MATRIX}
{SCIENCE}

The overview checkbox enables the overview table in the report. This can be useful for small searches, but becomes unwieldy for large searches. It is not very useful for a peptide mass fingerprint.

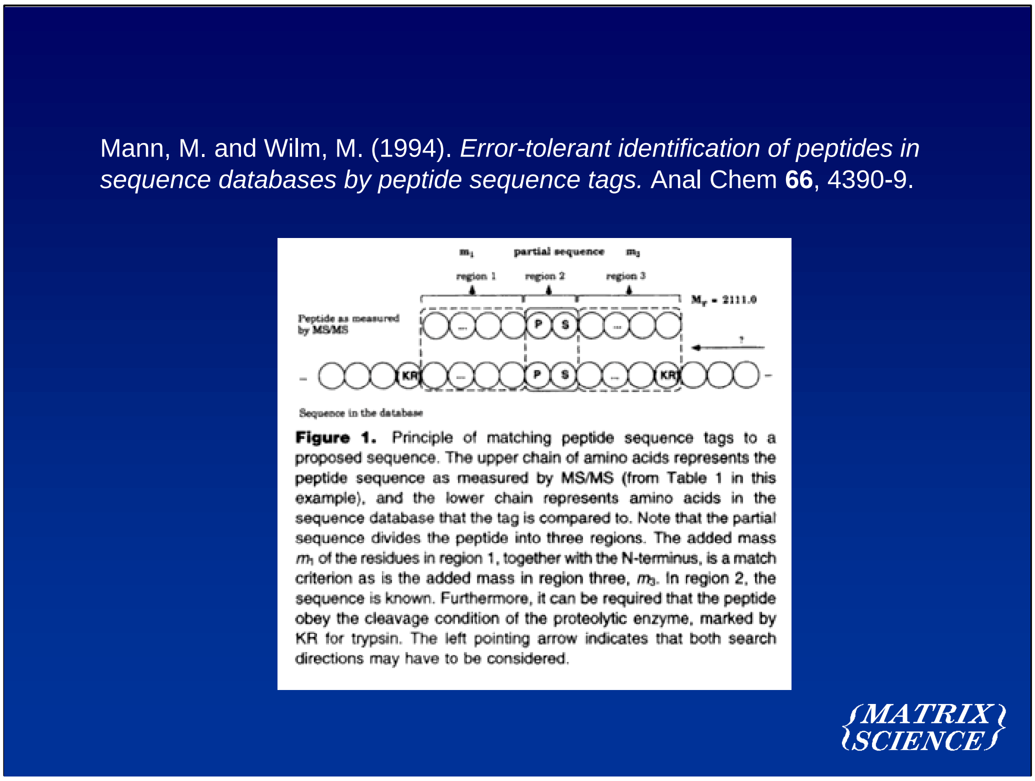


Three ways to use mass spectrometry data for protein ID:

1. **Peptide Mass Fingerprint**
   A set of peptide molecular weights from an enzyme digest of a protein
2. **Sequence Query**
   Mass values combined with amino acid sequence or composition data

The second method of protein identification using mass spectrometry data is the sequence query.

We define this as the general category of methods in which mass spectrometry data are combined with amino acid sequence or composition data.

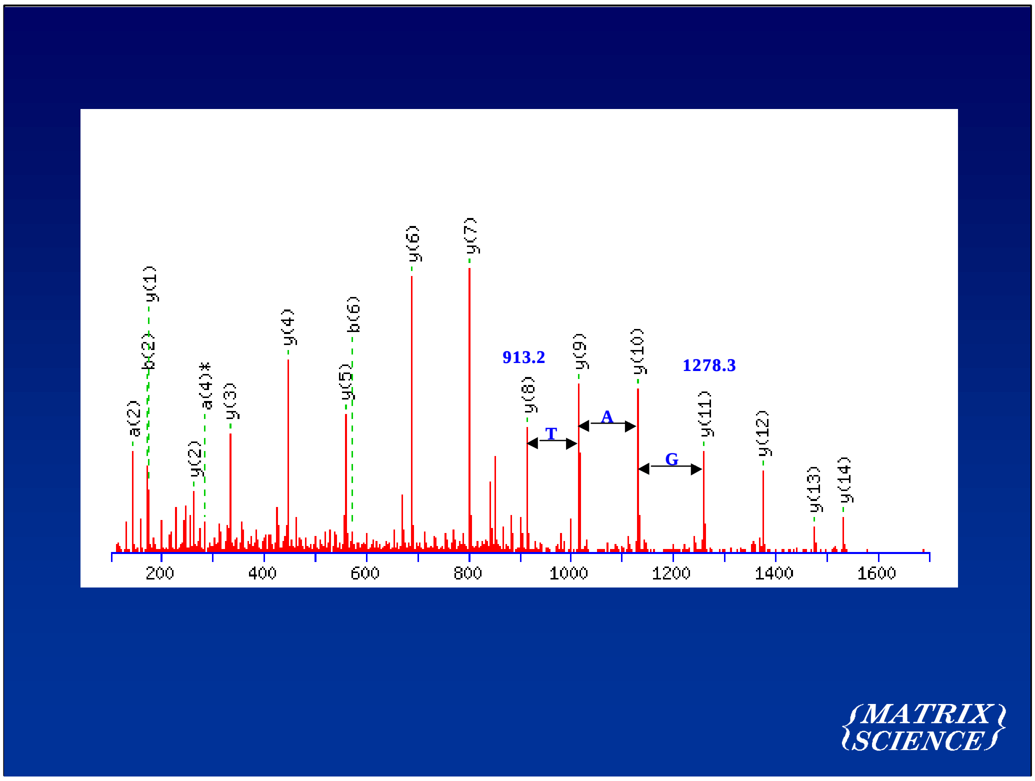


Mann, M. and Wilm, M. (1994). *Error-tolerant identification of peptides in sequence databases by peptide sequence tags.* Anal Chem **66**, 4390-9.

Figure 1. Principle of matching peptide sequence tags to a proposed sequence. The upper chain of amino acids represents the peptide sequence as measured by MS/MS (from Table 1 in this example), and the lower chain represents amino acids in the sequence database that the tag is compared to. Note that the partial sequence divides the peptide into three regions. The added mass m_1 of the residues in region 1, together with the N-terminus, is a match criterion as is the added mass in region three, m_3. In region 2, the sequence is known. Furthermore, it can be required that the peptide obey the cleavage condition of the proteolytic enzyme, marked by KR for trypsin. The left pointing arrow indicates that both search directions may have to be considered.

MATRIX SCIENCE

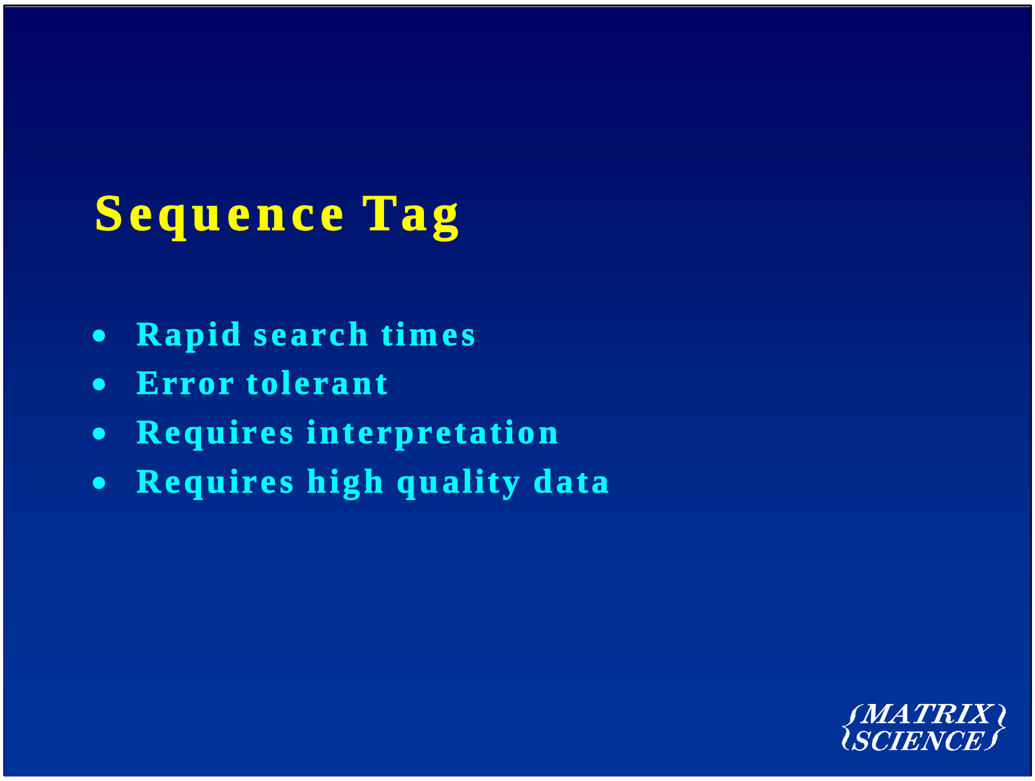
The most widely used approach in this category is the sequence tag approach, developed by Matthias Mann and colleagues at EMBL.



To perform a sequence tag search, it is necessary to manually interpret a few residues of amino acid sequence.

Even when the quality of the spectrum is poor, it is often possible to pick out four clean peaks, and read off three residues of sequence. In a sequence homology search, a triplet would be worth nothing. Any given triplet can be expected to occur by chance many times in even a small database.

What Mann and colleagues observed as that this very short stretch of amino acid sequence, when combined with the fragment ion mass values which enclose it and the peptide mass, could often provide sufficient specificity to provide an unambiguous identification.



Sequence Tag

- Rapid search times
- Error tolerant
- Requires interpretation
- Requires high quality data

MATRIX
SCIENCE

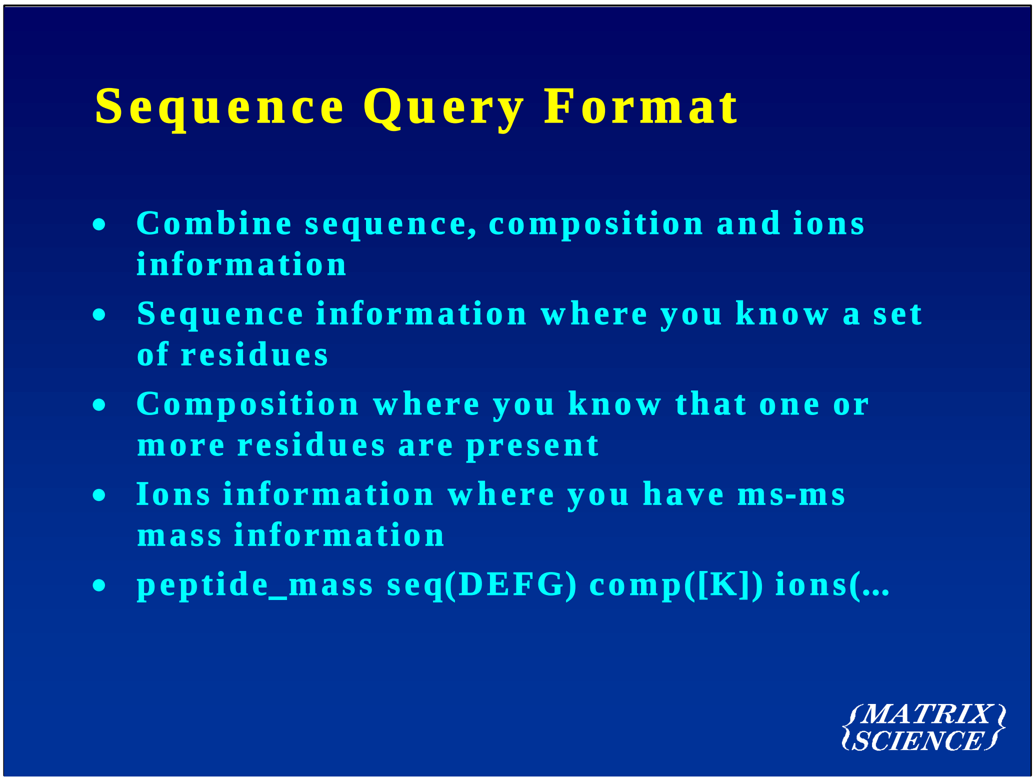
The two major advantages of the sequence tag approach are, first, that the search itself is fast, because it is basically just a filter on the database.

Second, this approach is amenable to error tolerant searching. What this means is that, even if the peptide contains (say) an amino acid substitution or an unknown post translational modification, it may still be possible to get a match by allowing the tag to “float”. That is, we look for the tag together with just one of the mass differences between the the tag and the peptide termini. This reduces the specificity of the tag, but it does allow for a mass difference to one side or the other.

On the down side, the sequence tag depends on interpretation of the data…sometimes by software, but more commonly by an expert user.

If the sequence is called wrongly, the search will fail. This means that high quality data are essential...good signal to noise and good mass accuracy.

Taken together, these last two features make the sequence tag approach less attractive for very high throughput work.



Sequence Query Format

- Combine sequence, composition and ions information
- Sequence information where you know a set of residues
- Composition where you know that one or more residues are present
- Ions information where you have ms-ms mass information
- peptide_mass seq(DEFG) comp([K]) ions(...

Using the Mascot sequence query, it is possible to combine sequence information, composition information and ms-ms mass values.

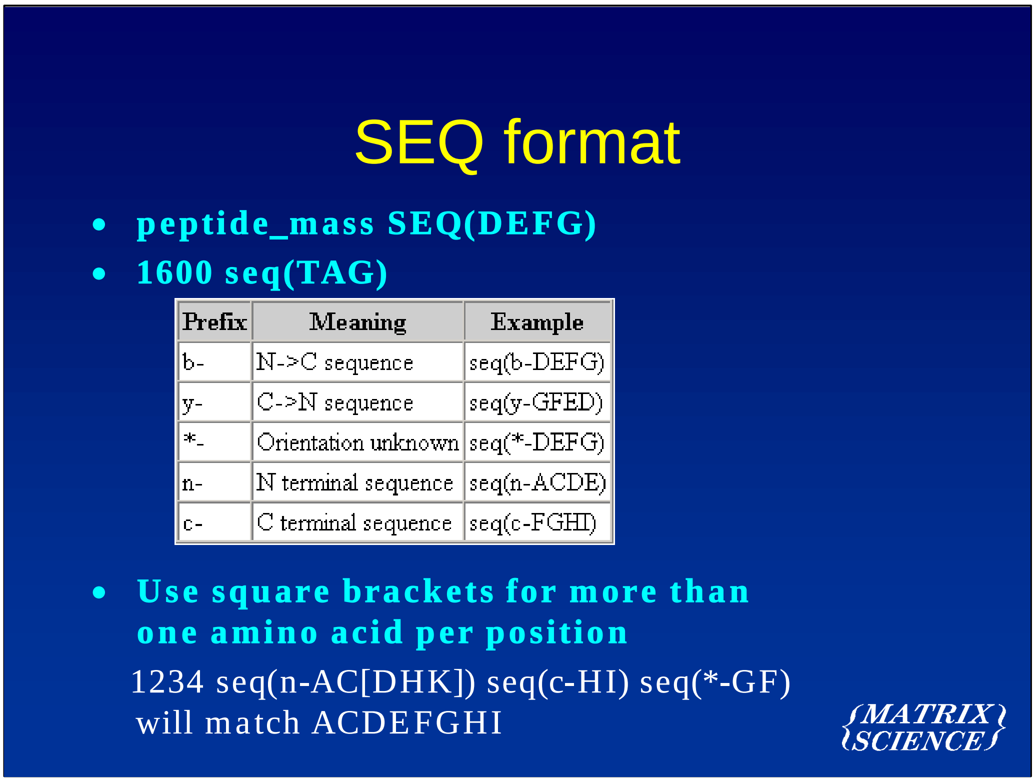
We have already seen sequence tags.

Composition information is used when you know that, for example there are exactly 3 methionines in the peptide.

MS-MS peak masses are generally entered using the MS-MS ions search form, but they can also be entered in the sequence query form.

It is possible to enter a mixture of all three different types of information for one peptide mass.

Now for some more detail on SEQ, COMP and IONS



SEQ format

- **peptide_mass SEQ(DEFG)**
- **1600 seq(TAG)**

Prefix	Meaning	Example
b-	N->C sequence	seq(b-DEFG)
y-	C->N sequence	seq(y-GFED)
*-	Orientation unknown	seq(*-DEFG)
n-	N terminal sequence	seq(n-ACDE)
c-	C terminal sequence	seq(c-FGHI)

- **Use square brackets for more than one amino acid per position**

1234 seq(n-AC[DHK]) seq(c-HI) seq(*-GF)
will match ACDEFGHI

The peptide mass always needs to be specified as the first item.

The text SEQ can be in upper or lower case:

One or more residues must be specified. In this case Mascot will search for a peptide with mass 1600 daltons that has a threonine, alanine and glycine anywhere in that peptide.

It is possible to specify more detailed information about the direction of the sequence, and whether the sequence is at a terminus by using a prefix. The b- and y- prefixes indicate that the specified residues can be anywhere in the peptide - they specify the direction.

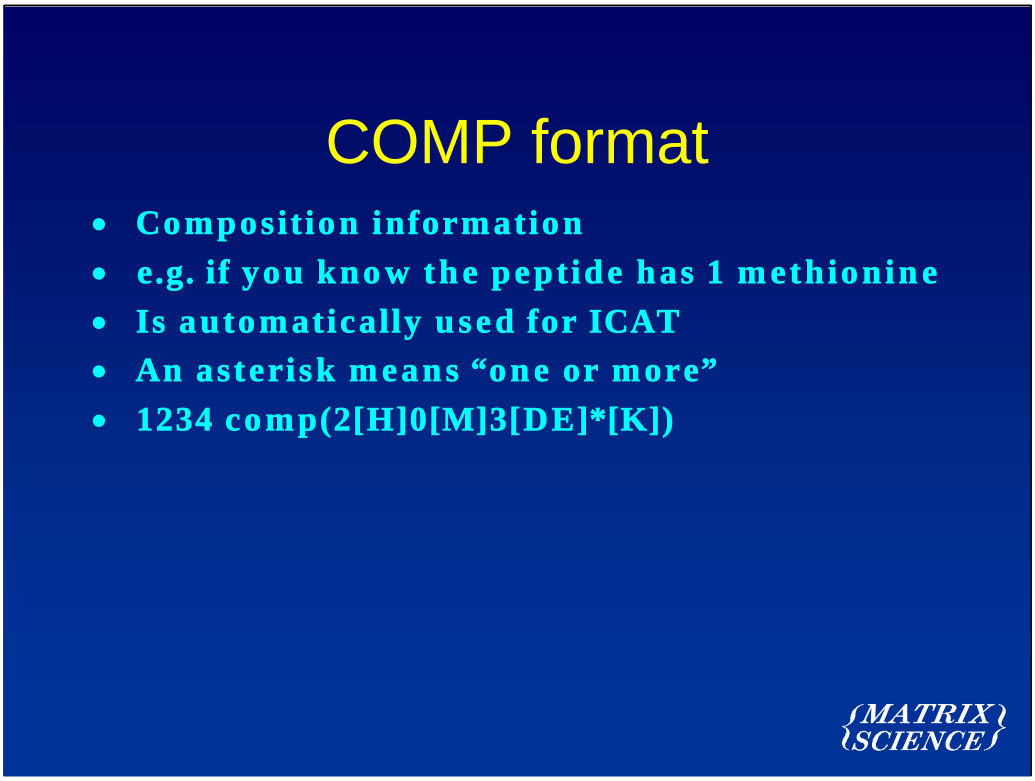
The n- and c- prefixes tie the sequence to the N and C terminus respectively. If no prefix is specified, then b- is assumed.

Use square brackets for when you have some doubt about a residue in a particular position.

It’s probably best to illustrate this with an example.

Peptide mass is 1234 - with the tolerance specified, and using any modifications specified.

The sequence starts with AC, and the next residue must be D,H or K. At the C terminus, there will be an HI. Somewhere else in the sequence, there will be an FG or a GF.



COMP format

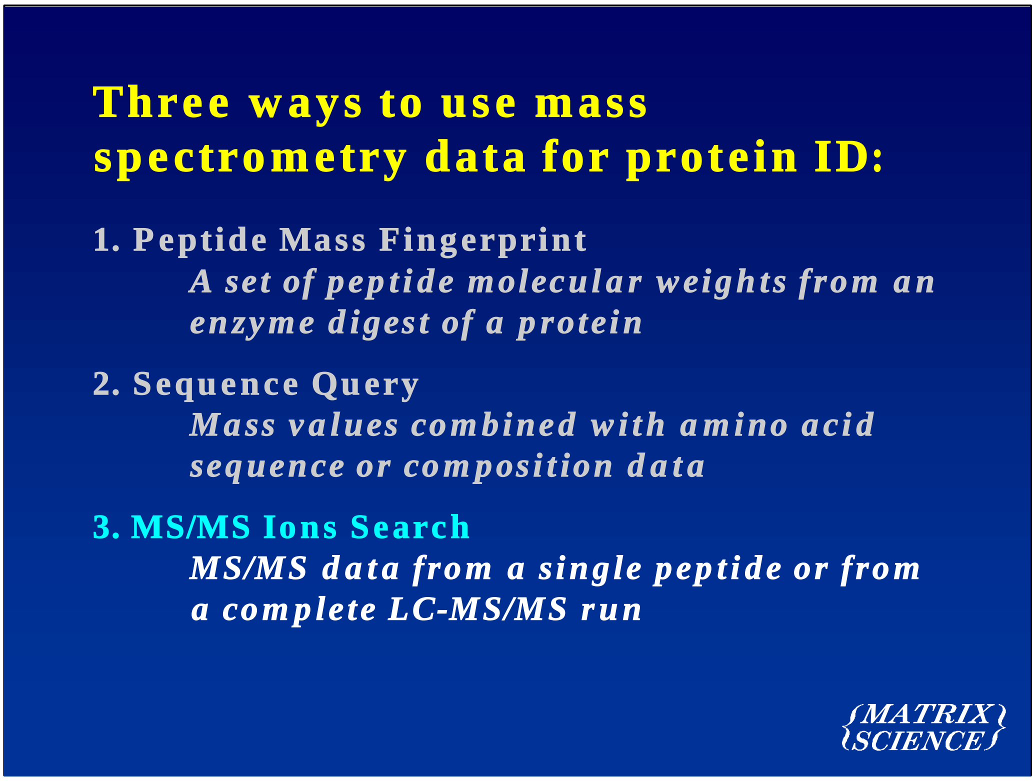
- Composition information
- e.g. if you know the peptide has 1 methionine
- Is automatically used for ICAT
- An asterisk means “one or more”
- 1234 comp(2[H]0[M]3[DE]*[K])

Comp is used to specify that certain residues are present or not present in a peptide. For example, you may know that a peptide has 1 or more methionines. If you don’t know exactly how many of a particular residue are correct, then an asterix can be used to indicate one or more.

A comp([C]*) command is inserted automatically in the latest version of Mascot when the ICAT button is selected, so that only peptides with one or more cysteine are searched.

In this example, a match would be found to a peptide of mass 1234 which contains 2 histidines, no methionines, 3 acidic residues (glutamic or aspartic acid) and at least 1 lysine.

Only one comp statement is allowed per query.

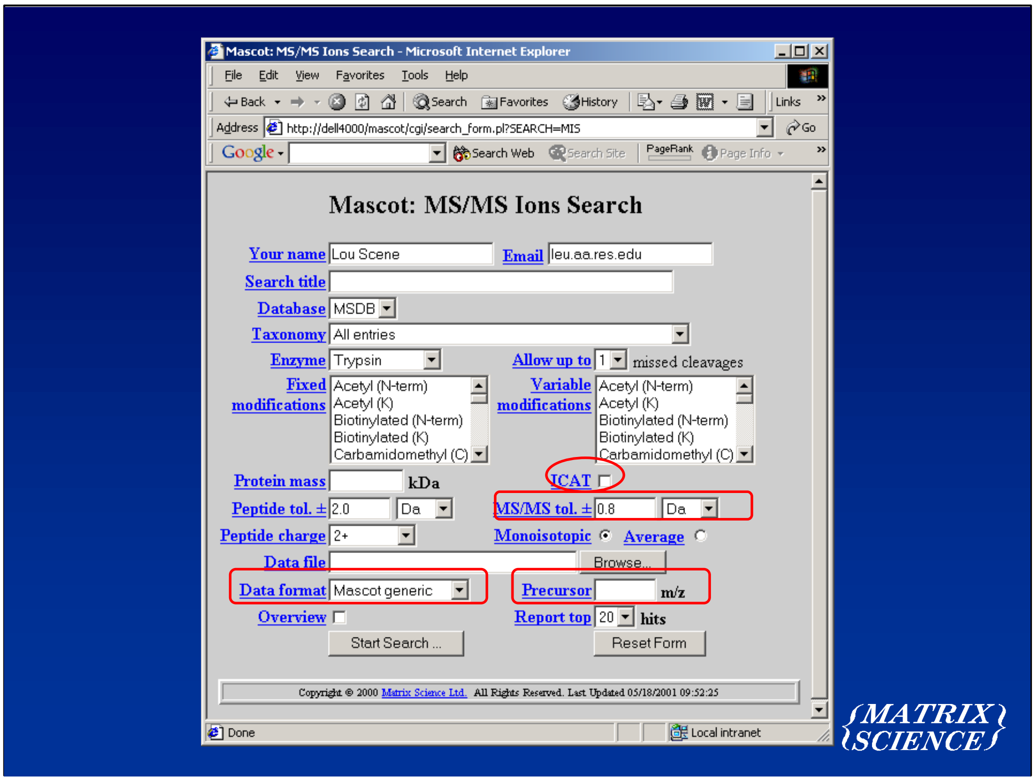


Three ways to use mass spectrometry data for protein ID:

1. **Peptide Mass Fingerprint**
   A set of peptide molecular weights from an enzyme digest of a protein
2. **Sequence Query**
   Mass values combined with amino acid sequence or composition data
3. **MS/MS Ions Search**
   MS/MS data from a single peptide or from a complete LC-MS/MS run

{MATRIX SCIENCE}

Which brings us to the third method of protein identification, searching uninterpreted MS/MS data. That is, using software to match lists of fragment ion mass and intensity values, without any manual sequence calling.


Mascot: MS/MS Ions Search - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites History Links
Google Search Web Search Site PageRank Page Info
Done
Local intranet

Address http://dell4000/mascot/cgi/search_form.pl?SEARCH=MIS

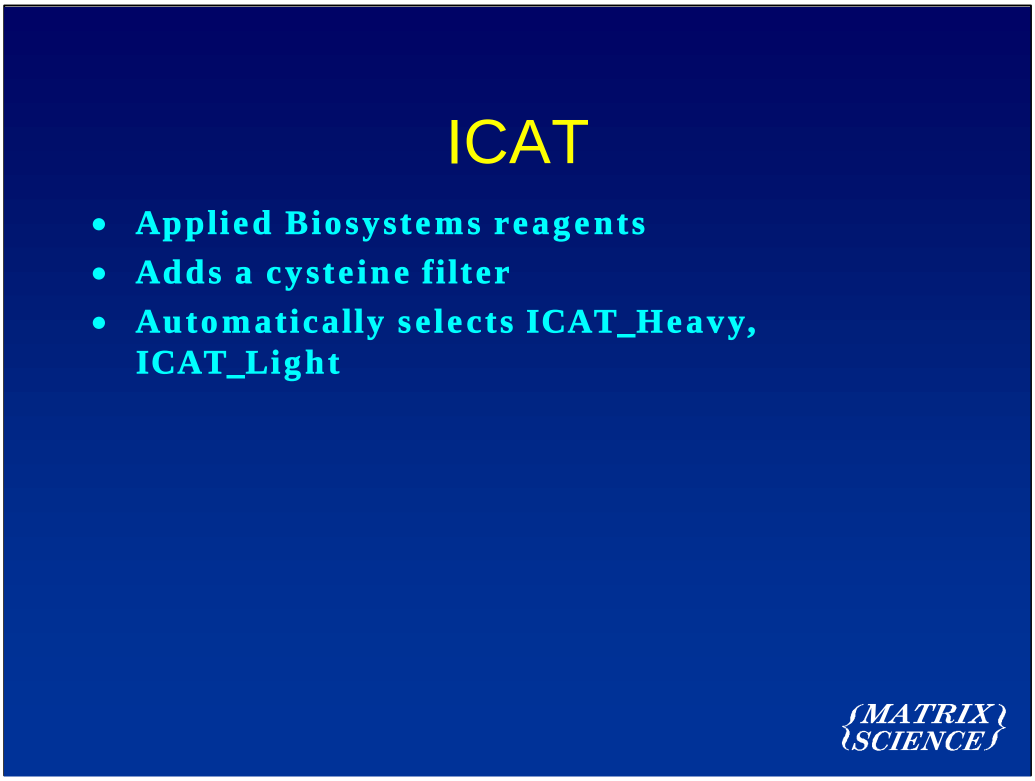
Mascot: MS/MS Ions Search

Your name: Lou Scene
Email: leu.aa.res.edu
Search title:
Database: MSDB
Taxonomy: All entries
Enzyme: Trypsin
Allow up to 1 missed cleavages
Fixed modifications: Acetyl (N-term), Acetyl (K), Biotinylated (N-term), Biotinylated (K), Carbamidomethyl (C)
Variable modifications: Acetyl (N-term), Acetyl (K), Biotinylated (N-term), Biotinylated (K), Carbamidomethyl (C)
Protein mass: kDa
ICAT □
Peptide tol. ± 2.0 Da
MS/MS tol. ± 0.8 Da
Peptide charge: 2+
Monoisotopic ● Average ○
Data file: Browse...
Data format: Mascot generic
Precursor: m/z
Overview □
Report top 20 hits
Start Search ... Reset Form

Copyright © 2000 Matrix Science Ltd. All Rights Reserved. Last Updated 05/18/2001 09:52:25

{MATRIX}
{SCIENCE}

The additional search parameters here are the ICAT button, the MS/MS tolerance parameter, the data file format, and the precursor m/z.



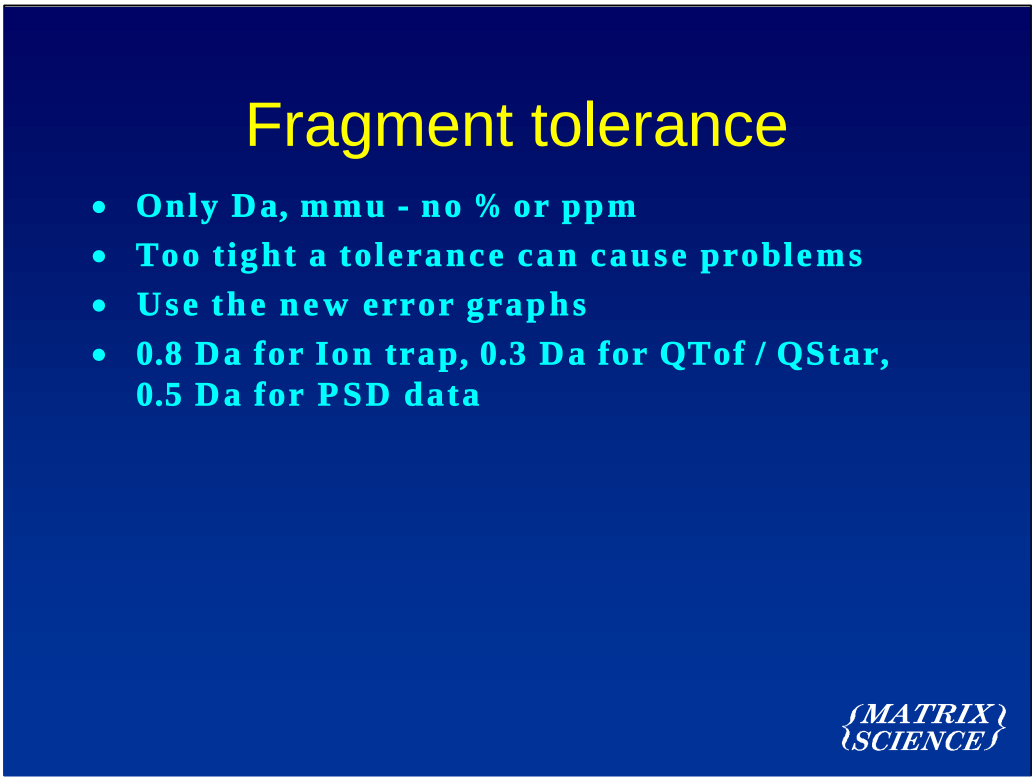
ICAT

- Applied Biosystems reagents
- Adds a cysteine filter
- Automatically selects ICAT_Heavy, ICAT_Light

{MATRIX}
{SCIENCE}

For ICAT, the default modification settings are for the Applied Biosystems reagents.

Selecting the ICAT check box adds a cysteine composition filter and selects the ICAT_Heavy and ICAT_Light modifications.



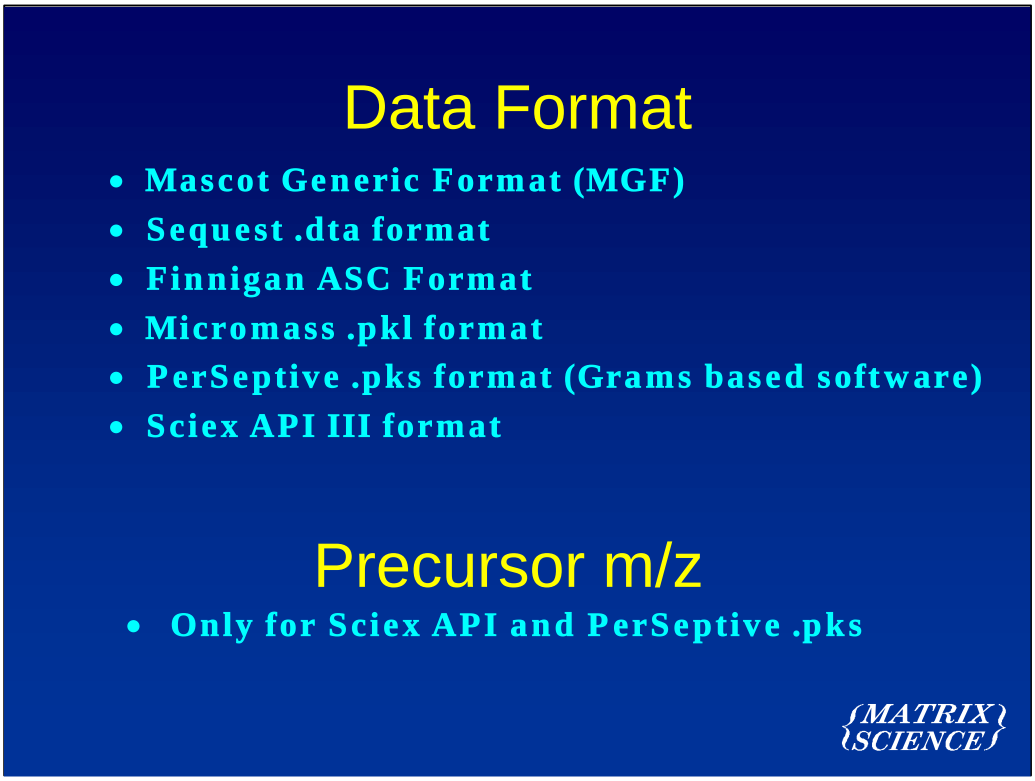
Fragment tolerance

- Only Da, mmu - no % or ppm
- Too tight a tolerance can cause problems
- Use the new error graphs
- 0.8 Da for Ion trap, 0.3 Da for QTof / QStar, 0.5 Da for PSD data

The fragment tolerance can only be entered in Daltons or milli-mass units - you cannot enter % or ppm.

Use the new error graphs in version 1.7 to help you determine sensible values for your instrument setup.

Once again, too tight a tolerance can cause problems.



Data Format

- Mascot Generic Format (MGF)
- Sequest .dta format
- Finnigan ASC Format
- Micromass .pkl format
- PerSeptive .pks format (Grams based software)
- Sciex API III format

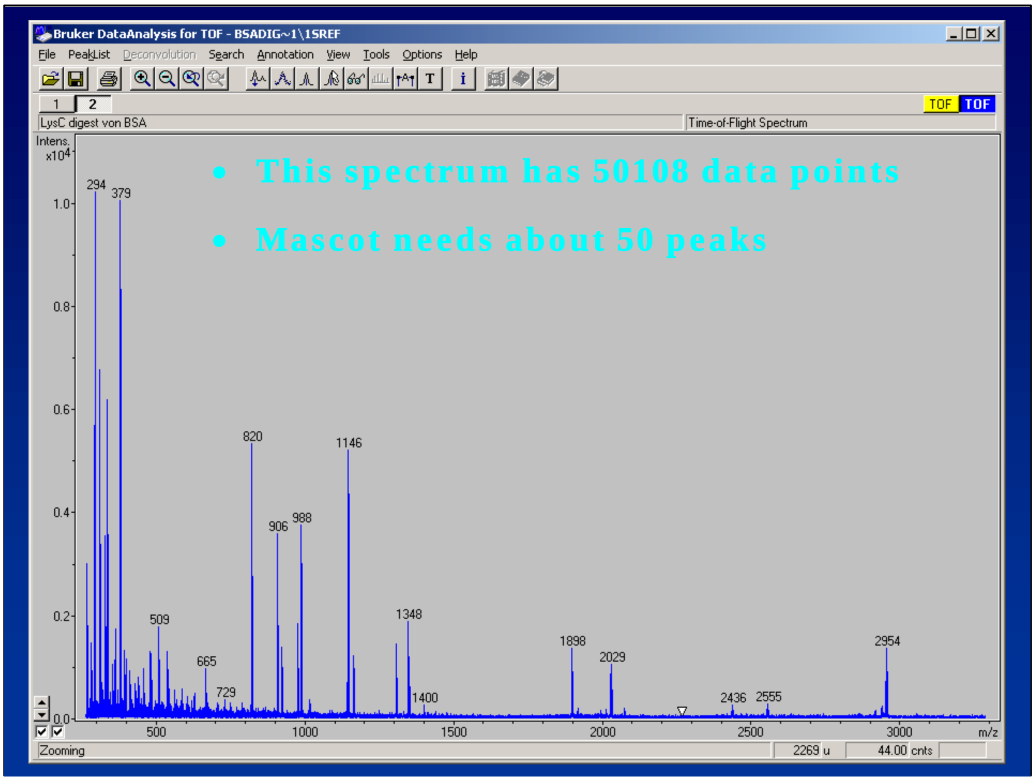
Precursor m/z

- Only for Sciex API and PerSeptive .pks

MATRIX SCIENCE

These MS-MS data formats are supported.

The precursor m/z value is only required for Sciex API and the old PerSeptive files. In these cases, note that only a single MS-MS spectrum can be submitted.



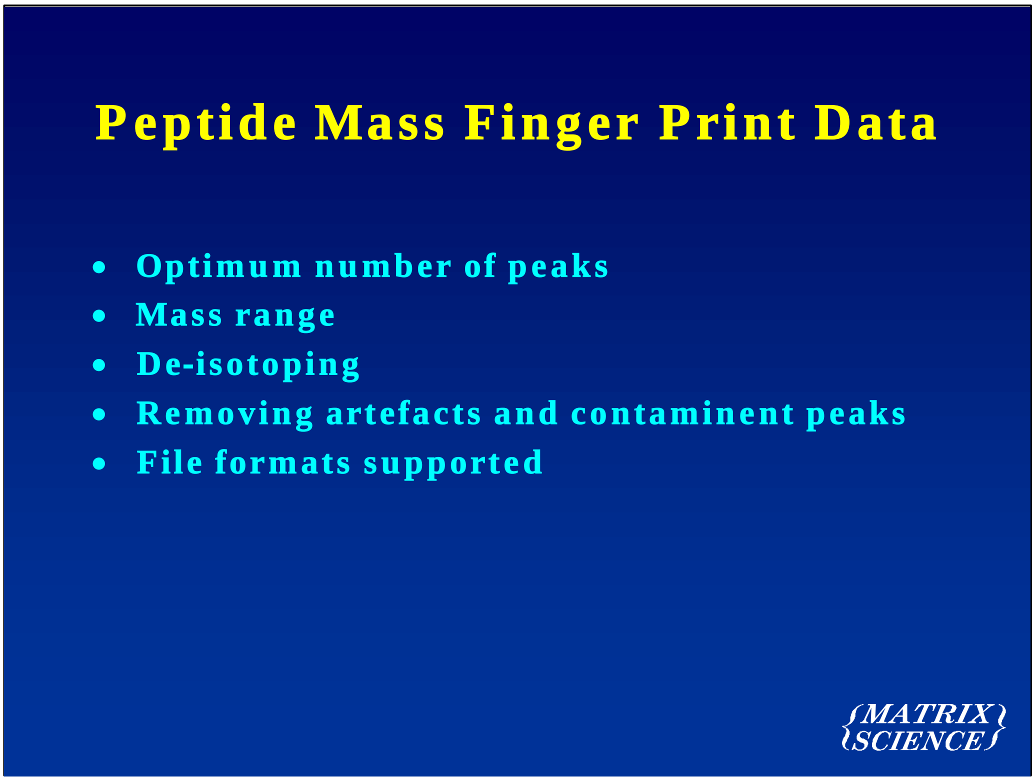
- This spectrum has 50108 data points
- Mascot needs about 50 peaks

Let's look at the peptide mass fingerprint data.

Mascot does not perform peak detection, so we cannot simply send 50,108 data points of profile data to Mascot.

Ideally we want to just detect all the ‘real’ peaks from this spectrum and submit these to Mascot.

The results that you get from Mascot are totally dependent upon the quality of the peak lists. It is all too easy to get a poor peak list, even from a great spectrum.



Peptide Mass Finger Print Data

- Optimum number of peaks
- Mass range
- De-isotoping
- Removing artefacts and contaminent peaks
- File formats supported

These are the main issues that we need to address with peptide mass fingerprint data:

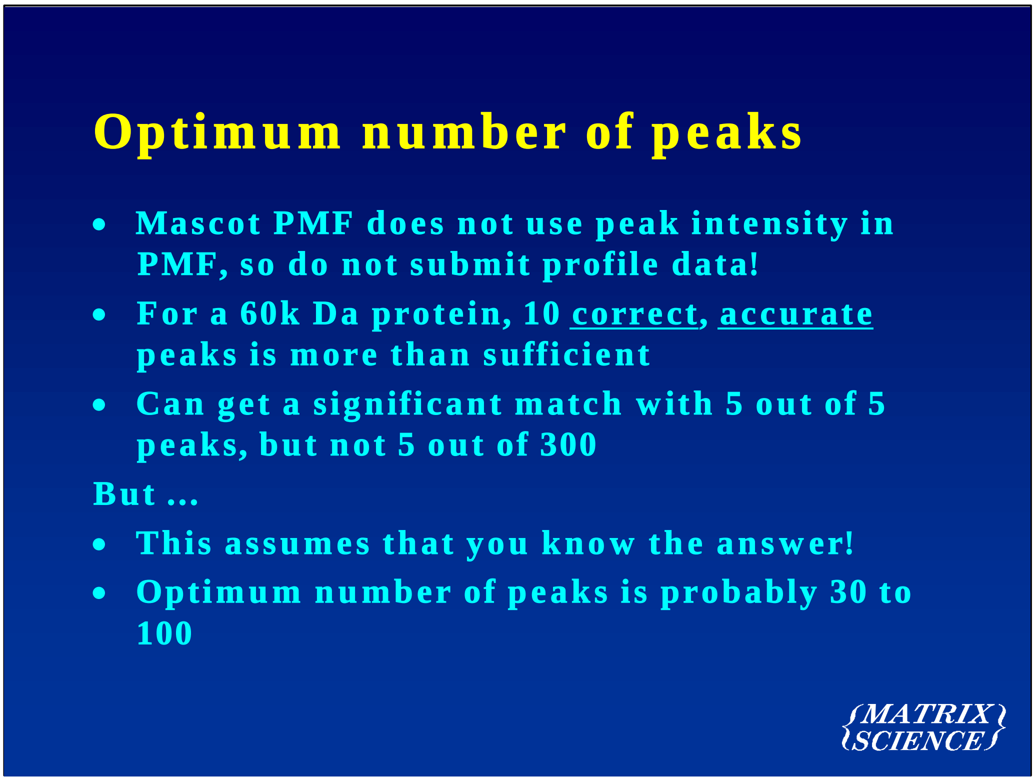
Optimum number of peaks

The mass range - e.g. is it worth submitting peaks with mass 100 Da?

Whether or not you de-isotope the data

Whether or not to remove contaminant peaks

What file formats are supported



Optimum number of peaks

- Mascot PMF does not use peak intensity in PMF, so do not submit profile data!
- For a 60k Da protein, 10 correct, accurate peaks is more than sufficient
- Can get a significant match with 5 out of 5 peaks, but not 5 out of 300

But …

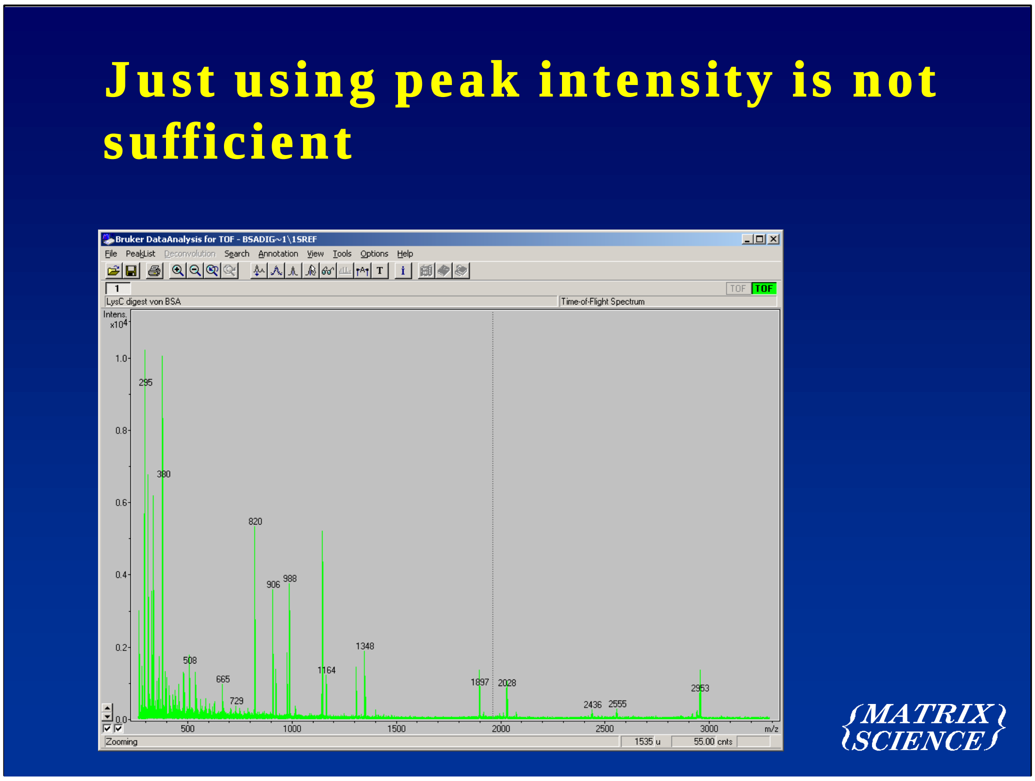
- This assumes that you know the answer!
- Optimum number of peaks is probably 30 to 100

{MATRIX}
{SCIENCE}

Mascot does not perform any peak detection with peptide mass fingerprint data. Each peak is equally important, so if you submit 300 peaks, then 300 peaks will be used. For each peak that fails to match, the Mascot score will be reduced, so it is important not to submit profile data or lists full of noise peaks.

If your data come from a 60kD protein, and you can supply 10 correct and accurate peaks, you will have a wonderful match. However, this assumes that you know what the answer is before you start. In reality, you will always have some peaks that don't match.

In general, you want to make sure that you try and submit all peaks with a good peak shape, so a good number of peaks is somewhere between 30 and 100 peaks.



Just using peak intensity is not sufficient

Bruker DataAnalysis for TOF - BSADIG~1\1SREF

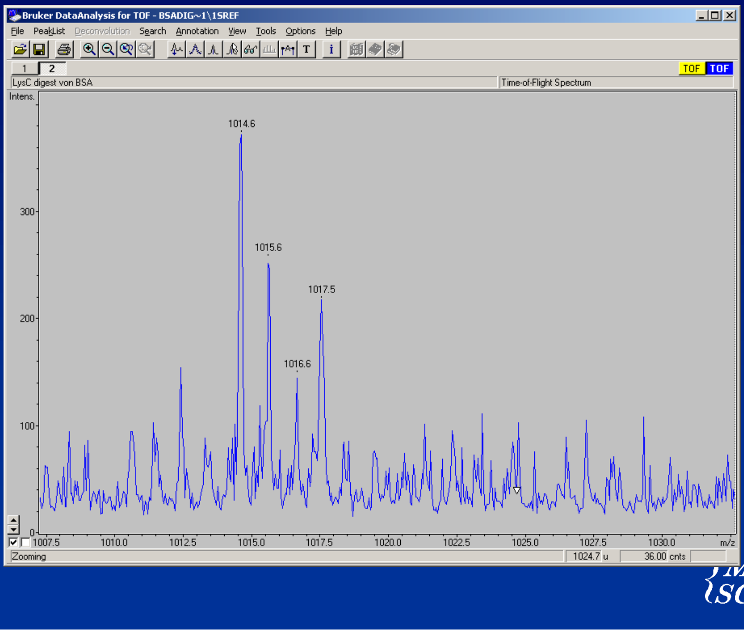
LysC digest von BSA

Time-of-Flight Spectrum

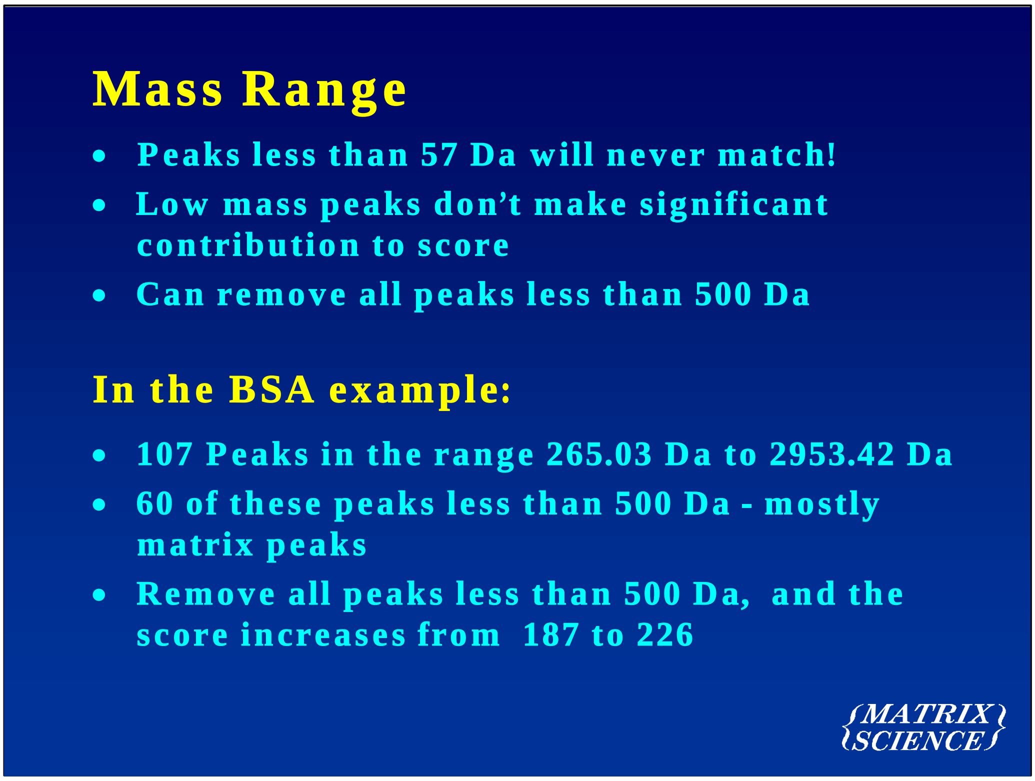
It is not sufficient to just pick the most intense peaks. From this spectrum, you probably can’t see that there is a reasonably good peak just above the noise level at about 1000 Daltons

Zooming in…

Small peak cluster at 1014.6



…shows a reasonable isotopic peak cluster at 1014.6 Daltons - just above the noise level. These peaks are only 1/100th of the intensity of the most intense peaks, but are just as valid as the more intense peaks.



Mass Range

- Peaks less than 57 Da will never match!
- Low mass peaks don’t make significant contribution to score
- Can remove all peaks less than 500 Da

In the BSA example:

- 107 Peaks in the range 265.03 Da to 2953.42 Da
- 60 of these peaks less than 500 Da - mostly matrix peaks
- Remove all peaks less than 500 Da, and the score increases from 187 to 226

It is obviously not worth submitting peaks that are less than the mass of Glycine - the lowest mass amino acid which is 57 Daltons.

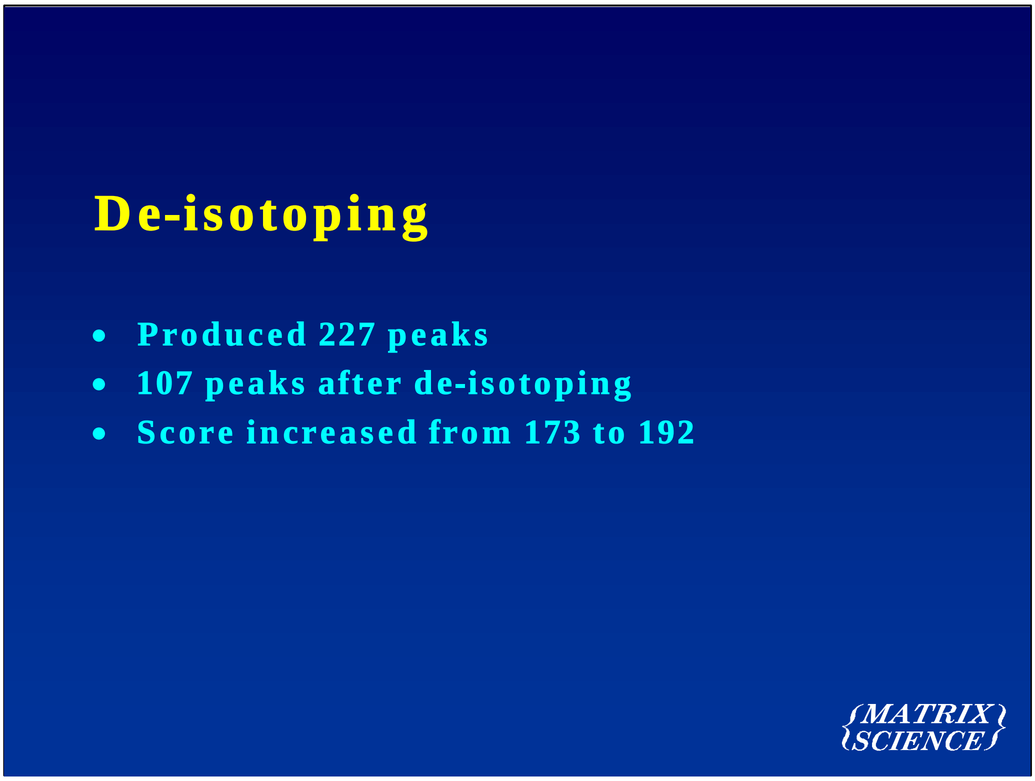
However, a single residue peptide is never going to add a significant amount to the score.

The same also applies to two, three and four residues peptides.

Also, consider that a MALDI spectrum will have many matrix peaks up to about 500 daltons, so in general, it is best to remove all masses below mass 500. This can often be a setting in the peak detection parameters.

Looking at the spectrum in the last slide, and using the default peak detection parameters across the whole range, there are 107 peaks detected. Of these, 60 are less than 500 daltons, and many are matrix peaks. By simply removing all the peaks less than 500 Daltons, the score increases from 187 to 226.

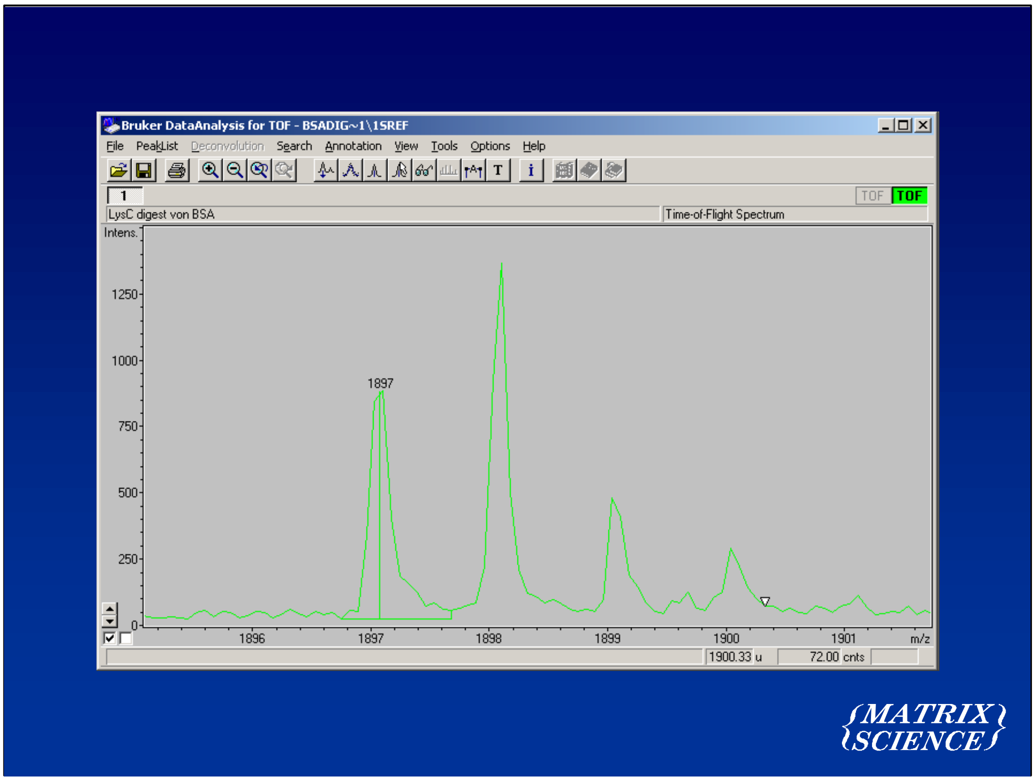
Just two small three residue peptides are lost, and these don’t have a significant effect on the score.



De-isotoping

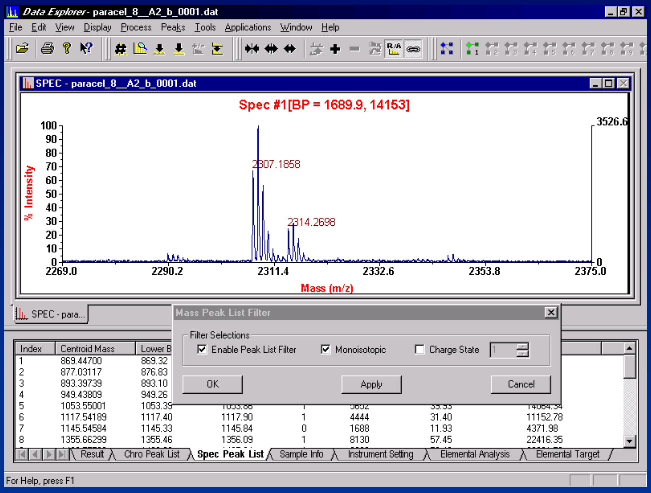
- Produced 227 peaks
- 107 peaks after de-isotoping
- Score increased from 173 to 192

The first pass detection on the BSA example spectrum using the default parameters produced 227 peaks. After de-isotoping, there were 107 peaks and the score increased from 173 to 192.



Most software packages include some de-isotoping routines - this is the Bruker Data Analysis software. In this example we have 4 peaks in an isotopic cluster, and after de-isotoping...

De-isotoping - data explorer



Mass Peak List Filter

Filter Selections

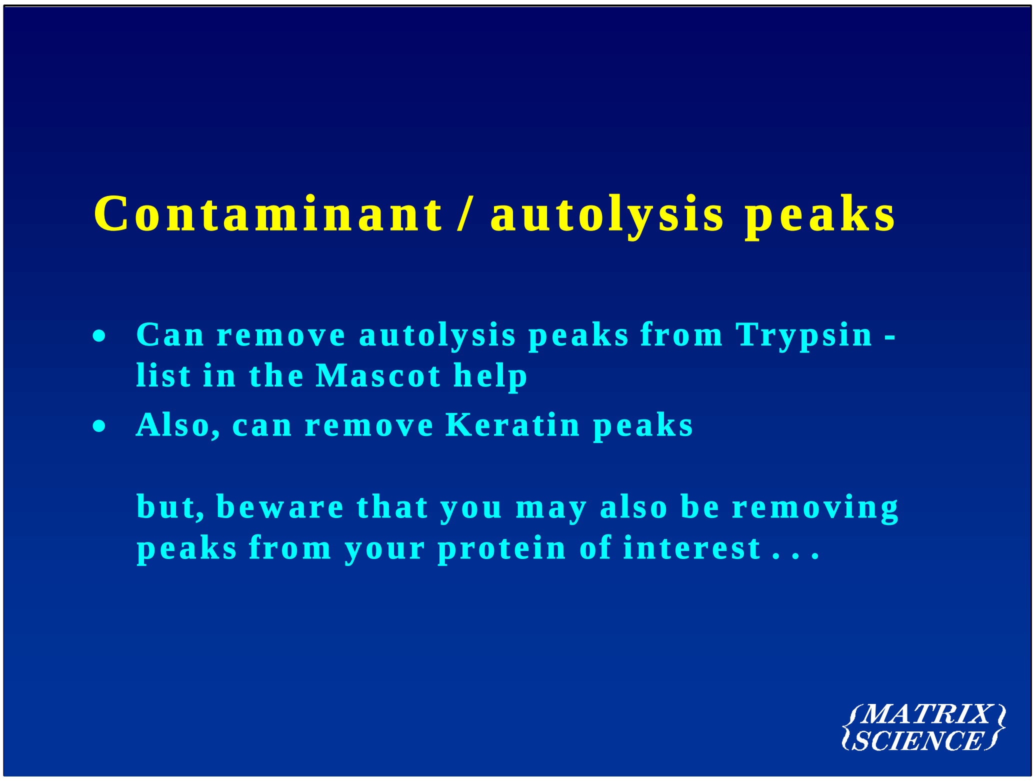
☑ Enable Peak List Filter ☑ Monoisotopic □ Charge State 1

OK Apply Cancel

Index	Centroid Mass	Lower B					
1	869.44700	869.32					
2	877.03117	876.83					
3	893.39739	893.10					
4	949.43809	949.26					
5	1053.55001	1053.39	1053.86	1	5632	39.53	14064.34
6	1117.54189	1117.40	1117.90	1	4444	31.40	11152.78
7	1145.54584	1145.33	1145.84	0	1688	11.93	4371.98
8	1355.66299	1355.46	1356.09	1	8130	57.45	22416.35

{MATRIX}
{SCIENCE}

It is impossible for us to go through the details of the peak detection parameters for each system today - another example is the Applied Biosystems Data Explorer software which has a de-isotoping option as you can see here...

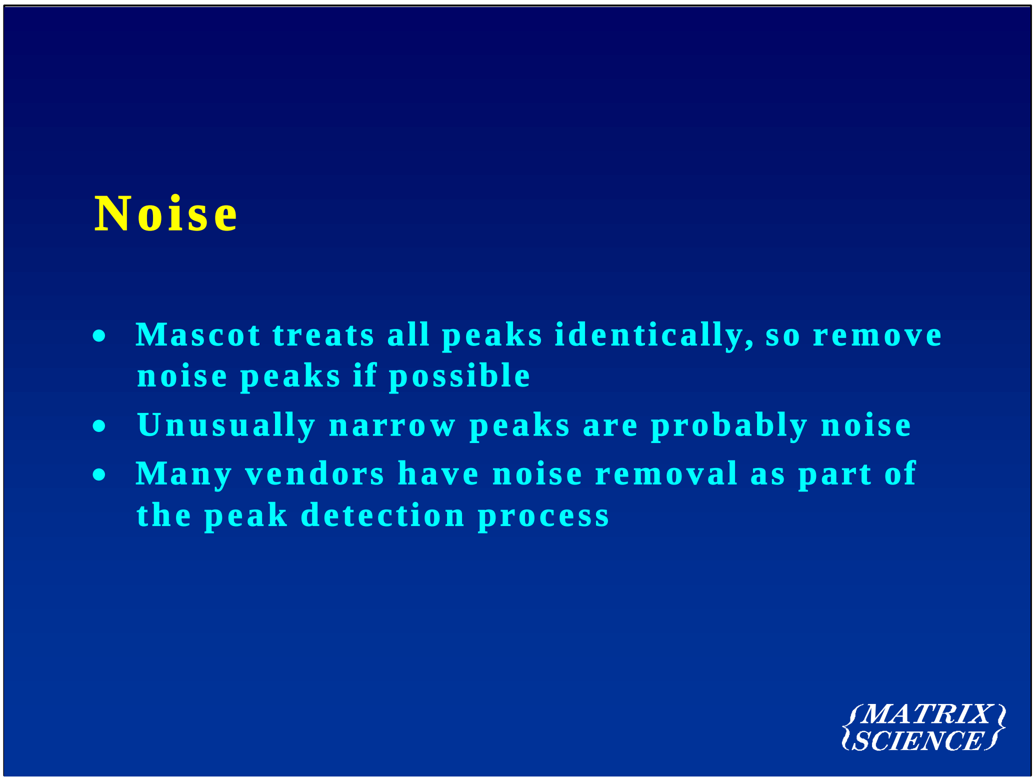


Contaminant / autolysis peaks

- Can remove autolysis peaks from Trypsin - list in the Mascot help
- Also, can remove Keratin peaks

but, beware that you may also be removing peaks from your protein of interest . . .

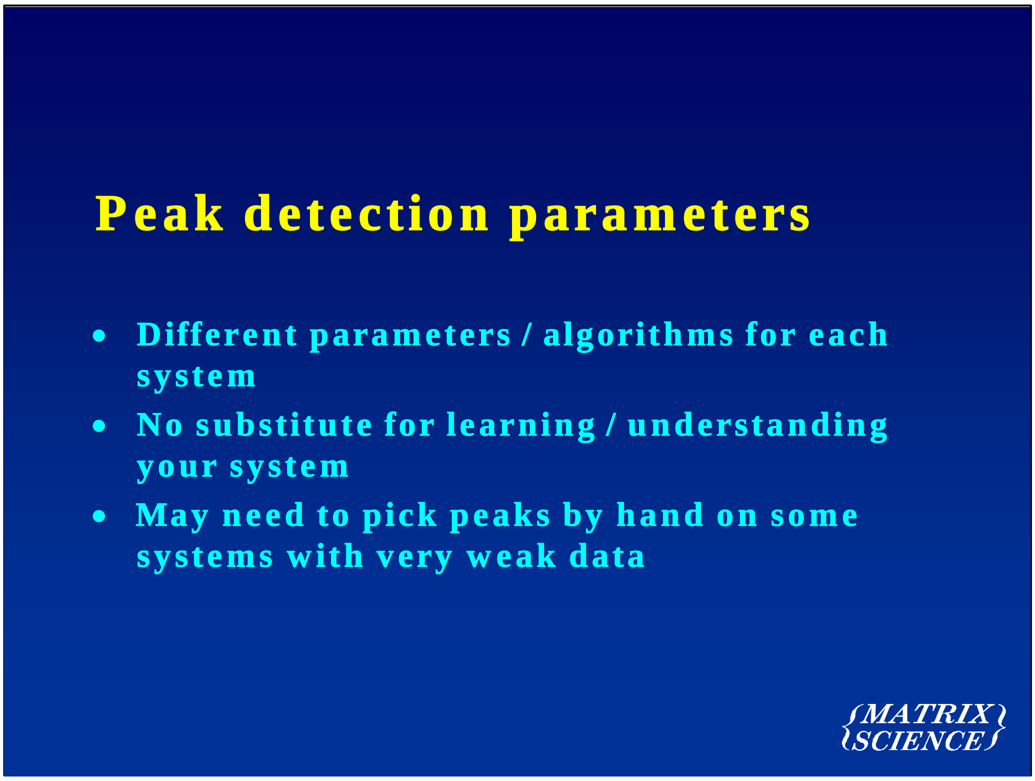
Some peaks may be produced by trypsin digesting itself. These can be removed before searching - however you should be aware that other peptides from your protein of interest may have identical masses, and so your results may suffer.



Noise

- Mascot treats all peaks identically, so remove noise peaks if possible
- Unusually narrow peaks are probably noise
- Many vendors have noise removal as part of the peak detection process

By the time the peak list reaches Mascot, there is no information about peak shape, so a noise peak will be treated identically to a peptide peak. If your data system includes functions to remove noise peaks, we recommend that you perform this step.

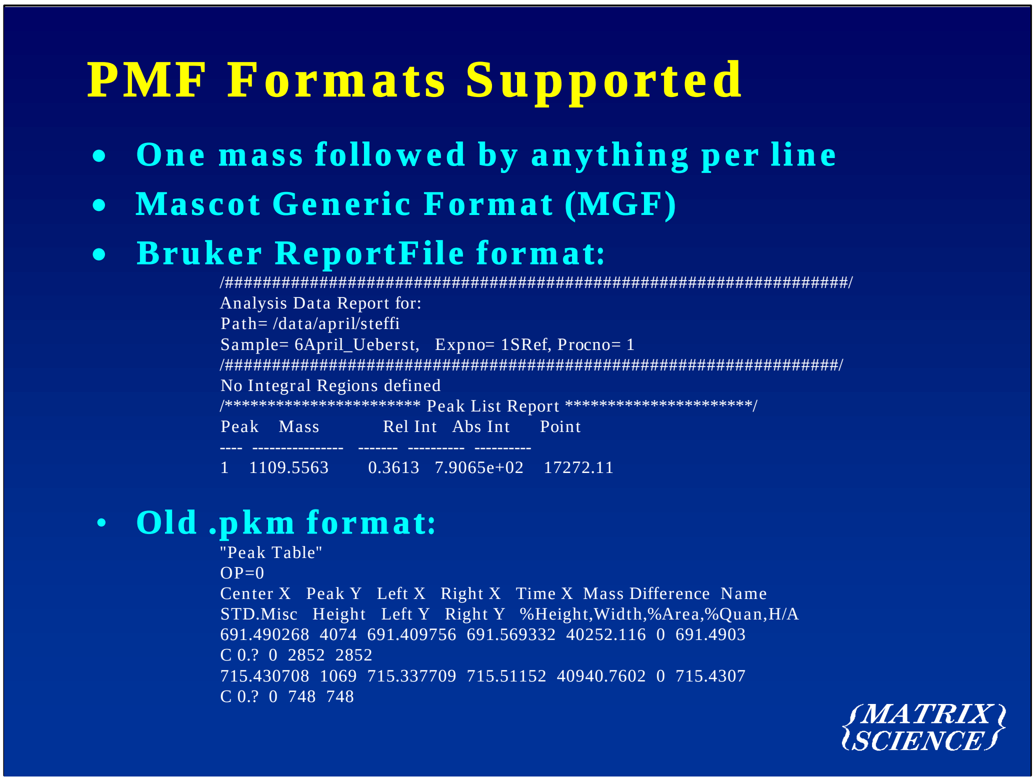


Peak detection parameters

- Different parameters / algorithms for each system
- No substitute for learning / understanding your system
- May need to pick peaks by hand on some systems with very weak data

It is impossible for us to go through the details of the peak detection parameters for the many different data systems today.

There is no substitute for understanding how your particular system works, and in the worst cases you may need to pick peaks by hand.



PMF Formats Supported

- One mass followed by anything per line
- Mascot Generic Format (MGF)
- Bruker ReportFile format:

```
/########################################################################/
Analysis Data Report for:
Path= /data/april/steffi
Sample= 6April_Ueberst,   Expno= 1SRef, Procno= 1
/#######################################################################/
No Integral Regions defined
/*********************** Peak List Report **********************/
Peak    Mass            Rel Int   Abs Int      Point
----  ---------------   -------  ----------  ----------
1    1109.5563        0.3613   7.9065e+02    17272.11
```

- Old .pkm format:

```
"Peak Table"
OP=0
Center X   Peak Y   Left X   Right X   Time X  Mass Difference  Name
STD.Misc   Height   Left Y   Right Y   %Height,Width,%Area,%Quan,H/A
691.490268  4074  691.409756  691.569332  40252.116  0  691.4903
C 0.?  0  2852  2852
715.430708  1069  715.337709  715.51152  40940.7602  0  715.4307
C 0.?  0  748  748
```

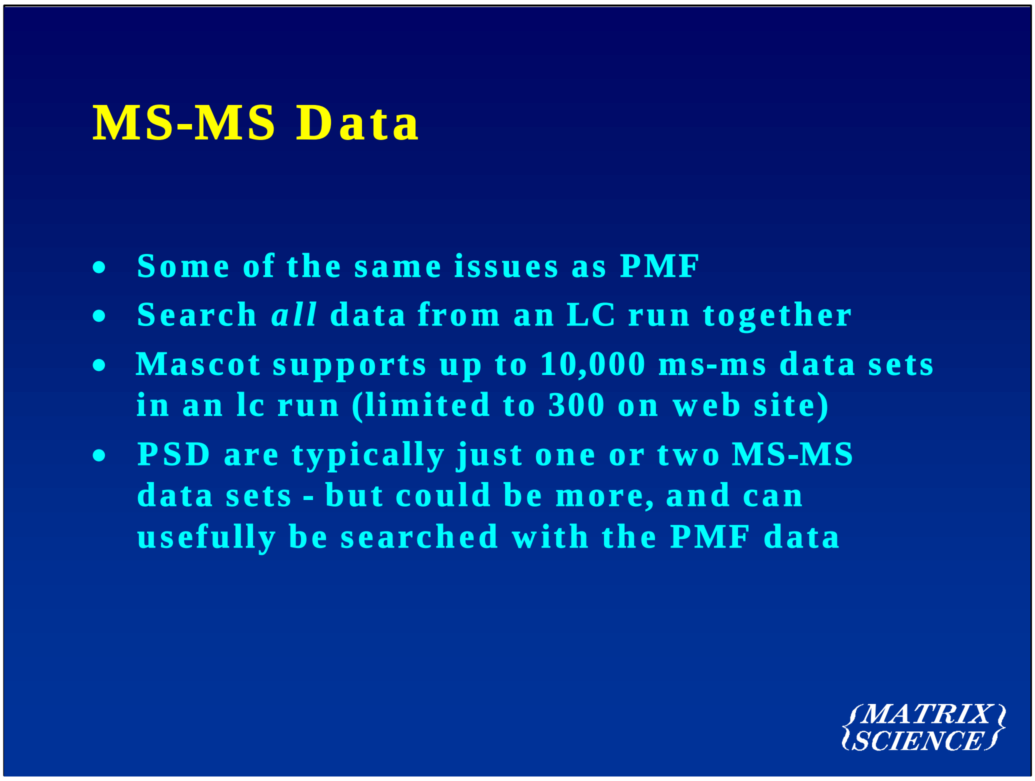
{MATRIX}
{SCIENCE}

The four main peptide mass fingerprint formats supported are:

A Mass value followed by anything per line. This covers most formats.

The Mascot Generic Format. This is described in the on-line help.

The Bruker ReportFile format and the PerSeptive Grams .pkm formats are supported as special cases.



MS-MS Data

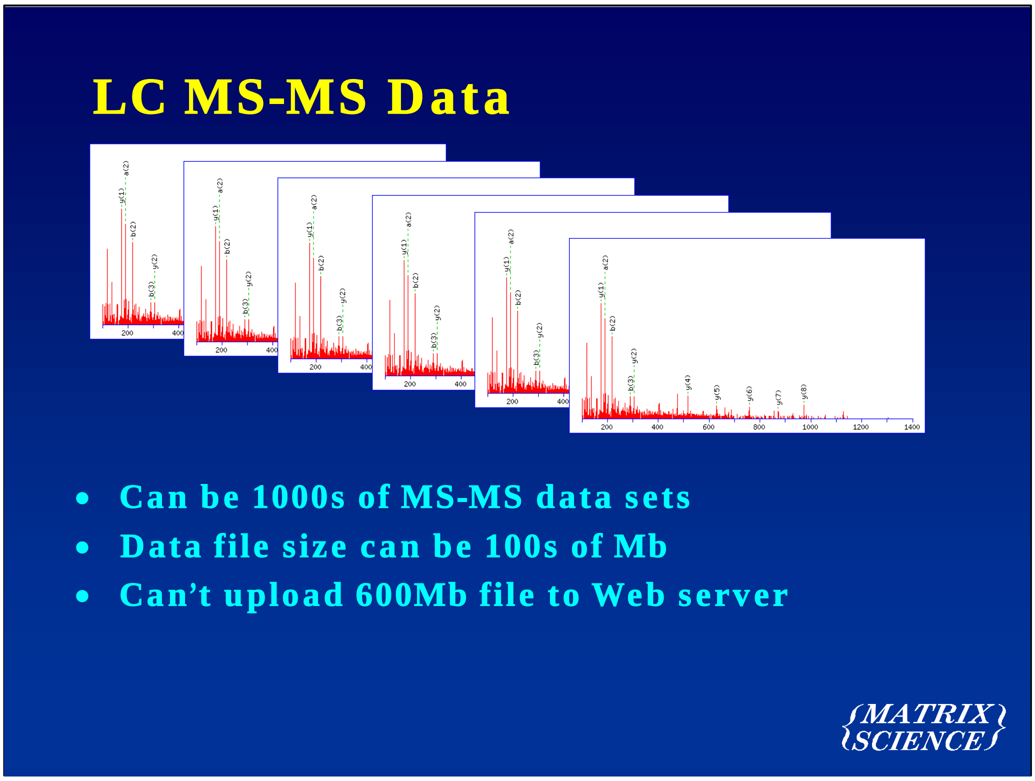
- Some of the same issues as PMF
- Search *all* data from an LC run together
- Mascot supports up to 10,000 ms-ms data sets in an lc run (limited to 300 on web site)
- PSD are typically just one or two MS-MS data sets - but could be more, and can usefully be searched with the PMF data

Moving on to MS-MS data, many of the issues are the same. However, the amount of data can be much greater.

It is normally best to search all data from an LC/MS-MS run together, so that multiple peptide matches from one protein will be grouped together.

Mascot supports up to 10,000 MS-MS spectra in a single search. On our public web site, we limit this to 300 spectra, to try and prevent the server from becoming overloaded.

PSD data are generally just one or two MS-MS data sets. These should also be searched together, possibly along with the PMF spectrum.



LC MS-MS Data

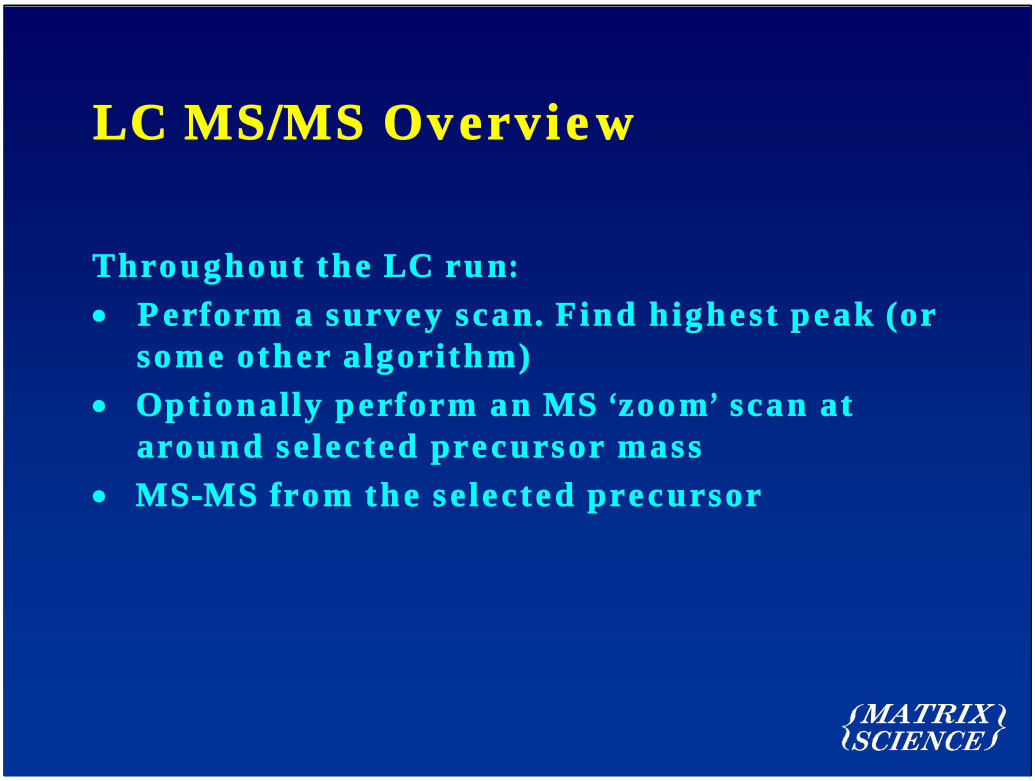
- Can be 1000s of MS-MS data sets
- Data file size can be 100s of Mb
- Can’t upload 600Mb file to Web server

MATRIX SCIENCE

For LC MS-MS data the problem is more acute. Typically, one MS-MS spectrum is being collected every second, and many people are performing several hour LC/MS runs.

This means that files can be be many 100s of Mb. We haven’t yet seen a file larger than 1Gb, but I guess that they have been created.

It isn’t practical to send say a 600Mb raw data file to the Matrix Science Web site (please don’t try!), so peak detection and some data reduction has to be performed.



LC MS/MS Overview

Throughout the LC run:

- Perform a survey scan. Find highest peak (or some other algorithm)
- Optionally perform an MS ‘zoom’ scan at around selected precursor mass
- MS-MS from the selected precursor

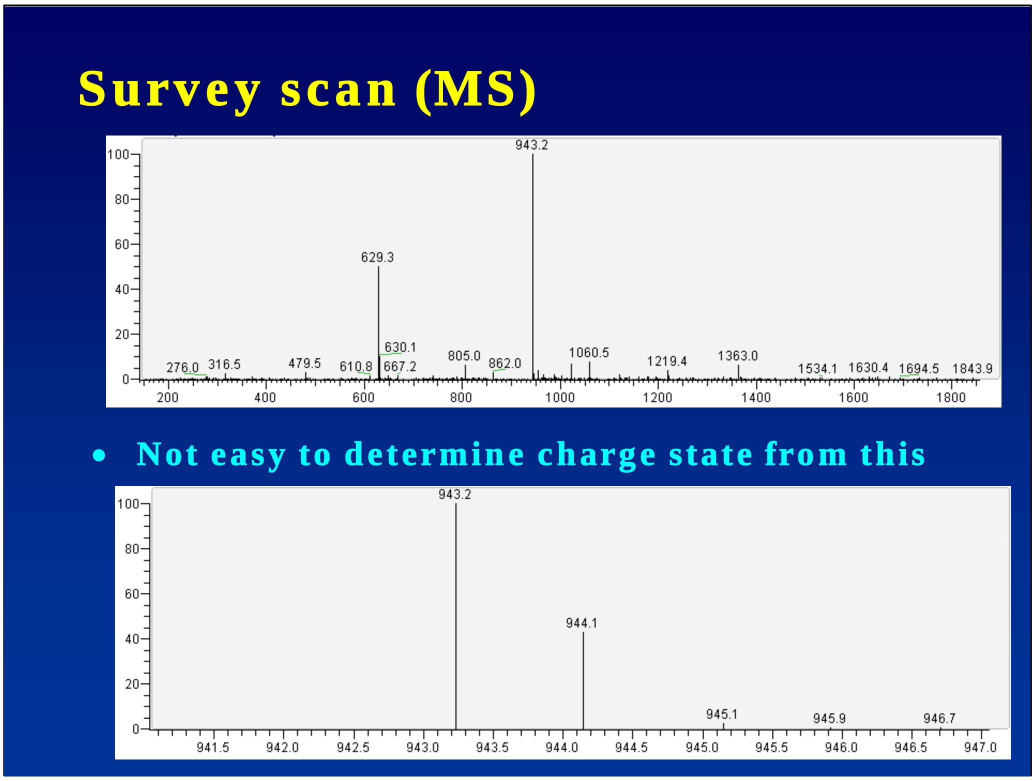
A quick overview of an LC run should be useful to get an understanding of what data is needed by Mascot. Some of this may be very obvious to many of you.

There are either two or three basic steps that are repeated throughout the run:

- A survey scan is performed. This spectrum is generally of no use for Mascot. Some technique is used to decide which peak or peaks should be singled out for MS-MS scan.

- On an ion trap, an MS scan is performed on a small mass range around the mass of the selected peptide mass. This is used to determine the mass more accurately, and enables the charge state to be determined.

- Finally the peptide is fragmented to produce an MS-MS scan. This is the only spectrum that needs to be searched by Mascot.

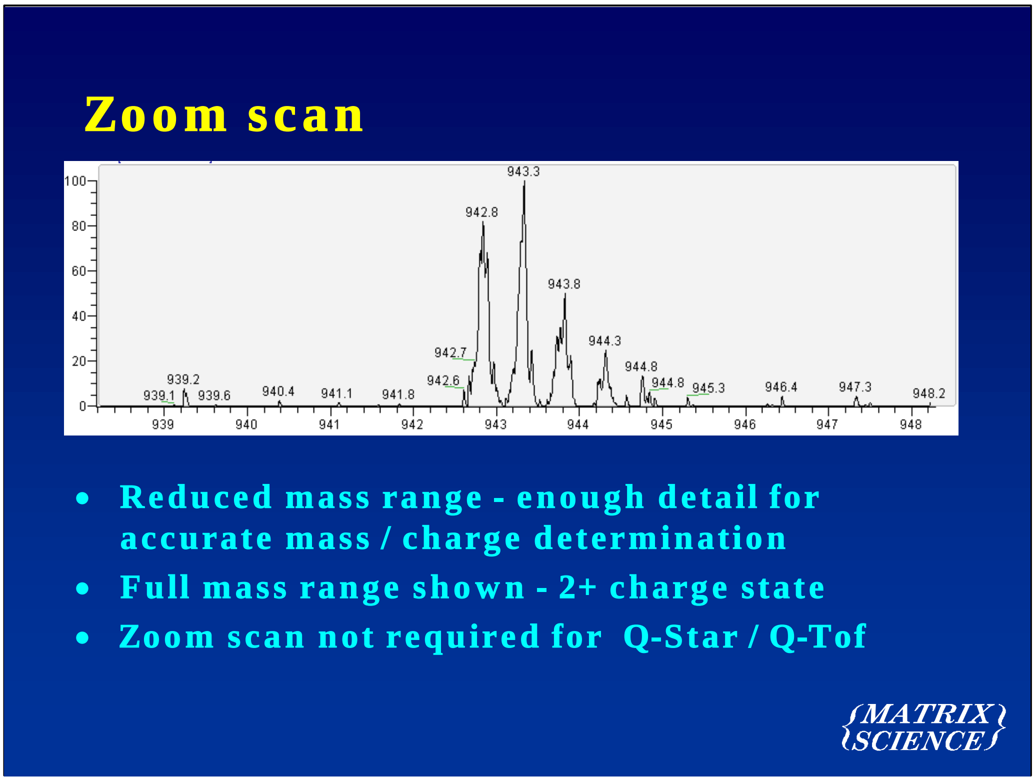


Survey scan (MS)

- Not easy to determine charge state from this

Here is an example of a survey scan, and the instrument software has chosen the most intense peak to fragment. If we just zoom in on the peak at 943.2, we can see that it is hard to determine the charge state from this.

You may think that this has a single charge state because the peaks appear to be almost a dalton apart - this is probably an artefact, because the data is being saved as centroided data.



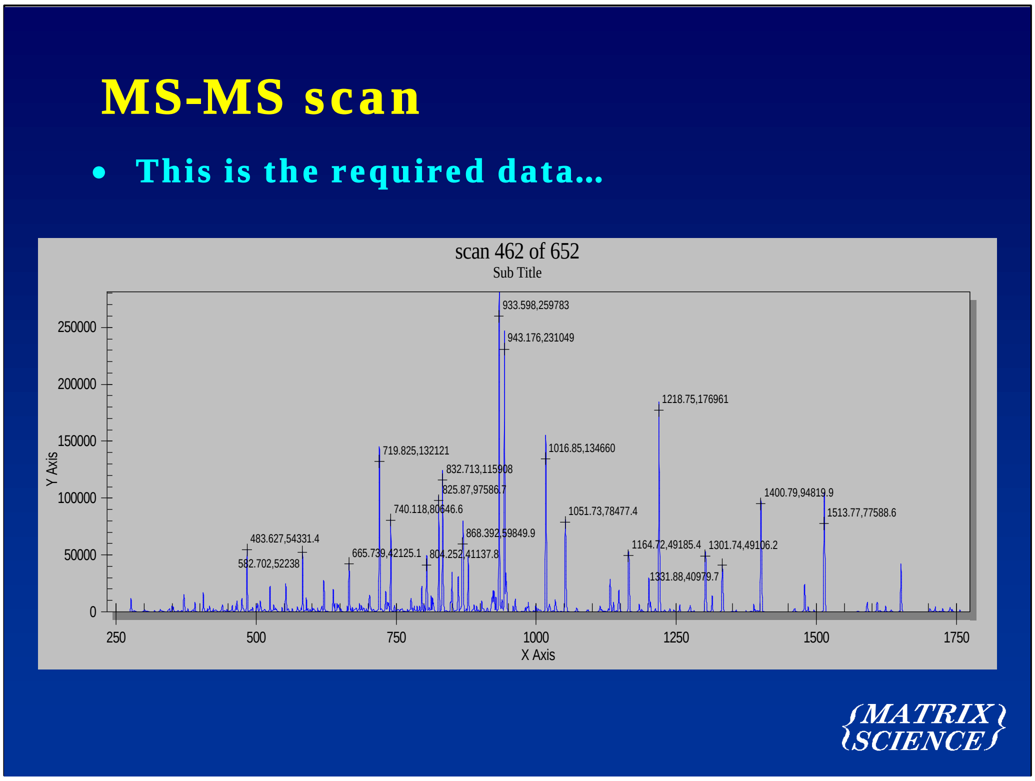
Zoom scan

- Reduced mass range - enough detail for accurate mass / charge determination
- Full mass range shown - 2+ charge state
- Zoom scan not required for Q-Star / Q-Tof

A zoom scan is performed on a reduced mass range, and we can see clearly that in this cluster of peaks, the peaks are half a dalton apart, so the charge state is 2.

Note that the full mass range is shown here - I have not zoomed in to get this spectrum

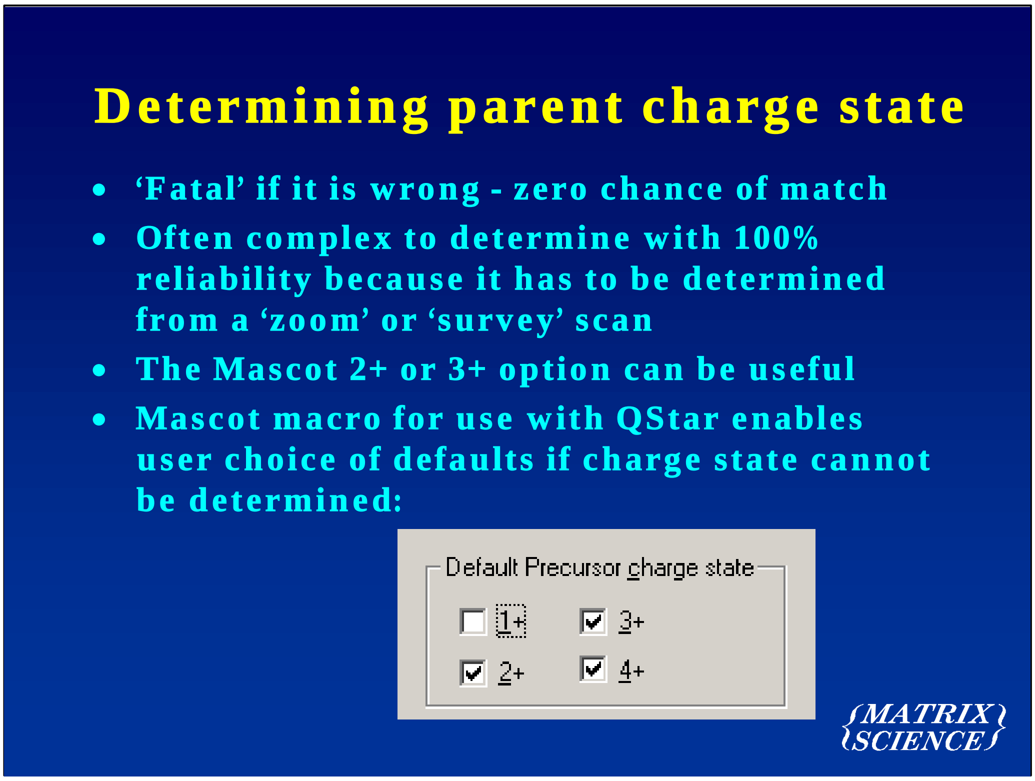
A zoom scan is not required on a Q-Star or Q-Tof instrument, because the resolution is higher.



MS-MS scan

- This is the required data…

So, we finally get to one of the many MS-MS spectra that we are interested in. This will either be saved as profile or centroided data by the instrument software.



Determining parent charge state

- ‘Fatal’ if it is wrong - zero chance of match
- Often complex to determine with 100% reliability because it has to be determined from a ‘zoom’ or ‘survey’ scan
- The Mascot 2+ or 3+ option can be useful
- Mascot macro for use with QStar enables user choice of defaults if charge state cannot be determined:

Default Precursor charge state

☐ 1+ ☑ 3+

☑ 2+ ☑ 4+

If the parent charge state is determined incorrectly, then Mascot will not get a match, because it will be filtering out the wrong peptides to search.

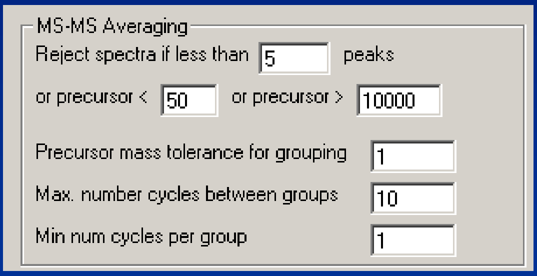
As we have seen, it is complex to determine the charge state, and things can go wrong.

Using the "2+ or 3+" option with Mascot for cases where the charge state is not determined properly can be useful.

With the Mascot macro for the Sciex software, it is possible to select a number of charge states for Mascot to try when it hasn't been possible to determine the charge state.

Averaging data from LC/MS-MS

- Multiple ms-ms spectra with same precursor
- These should be summed together before peak detection
- Typical parameters available for averaging:



MS-MS Averaging

Reject spectra if less than 5 peaks

or precursor < 50 or precursor > 10000

Precursor mass tolerance for grouping 1

Max. number cycles between groups 10

Min num cycles per group 1

MATRIX SCIENCE

The other major issue with LC/MS/MS runs is that multiple MS-MS spectra are obtained for the same precursor.

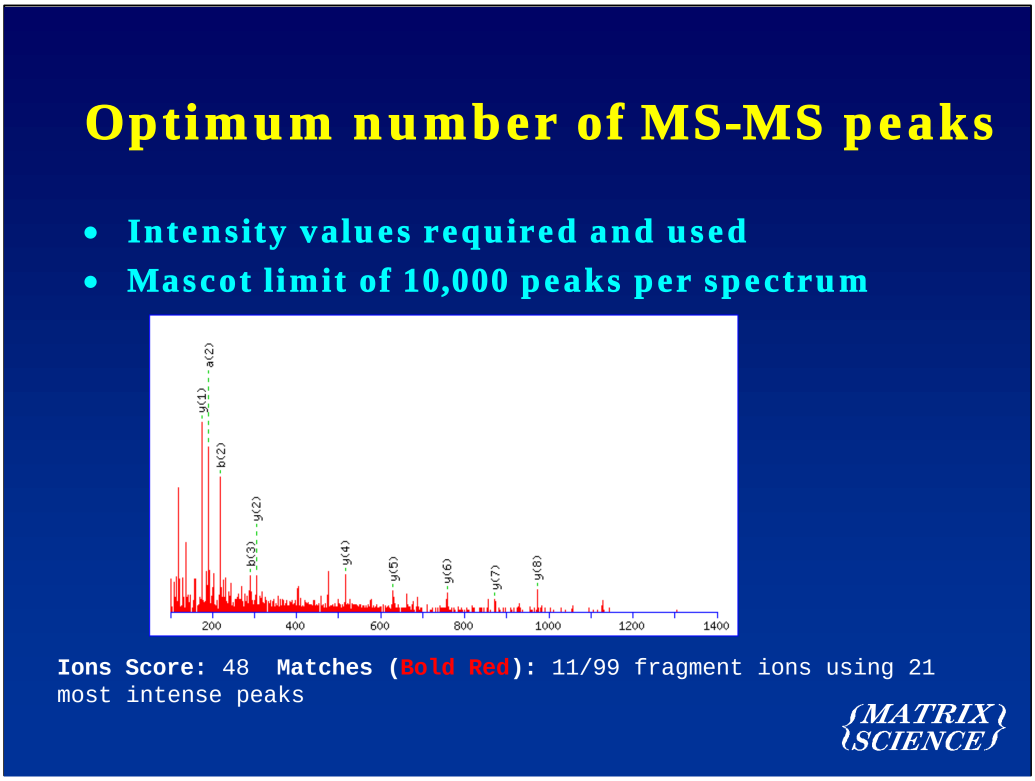
These need to be summed together before they are submitted to Mascot. Typical parameters for this are shown here:

It is just not possible to get a significant match from an MS/MS spectrum with less than about 5 or 10 peaks - use this parameter filters out spectra with just a few noise spikes. Also, a very low or very high parent mass is not going to give good results.

To group spectra together requires that they have the same precursor mass - of course they won't be identical, so you need to specify a tolerance. For an ion trap, a value of 1 or 2 Daltons is probably correct.

If a peptide of mass 1000 is detected at the start of an LC run, and 20 minutes later another peptide of the same mass is detected, then these are probably not the same peptide assuming that the chromatography is functioning correctly. The max. number of cycles between groups allows you to prevent these two scans from being grouped together.

On some systems, a decent spectrum will only be obtained if you sum together more than one MS-MS spectrum.



Optimum number of MS-MS peaks

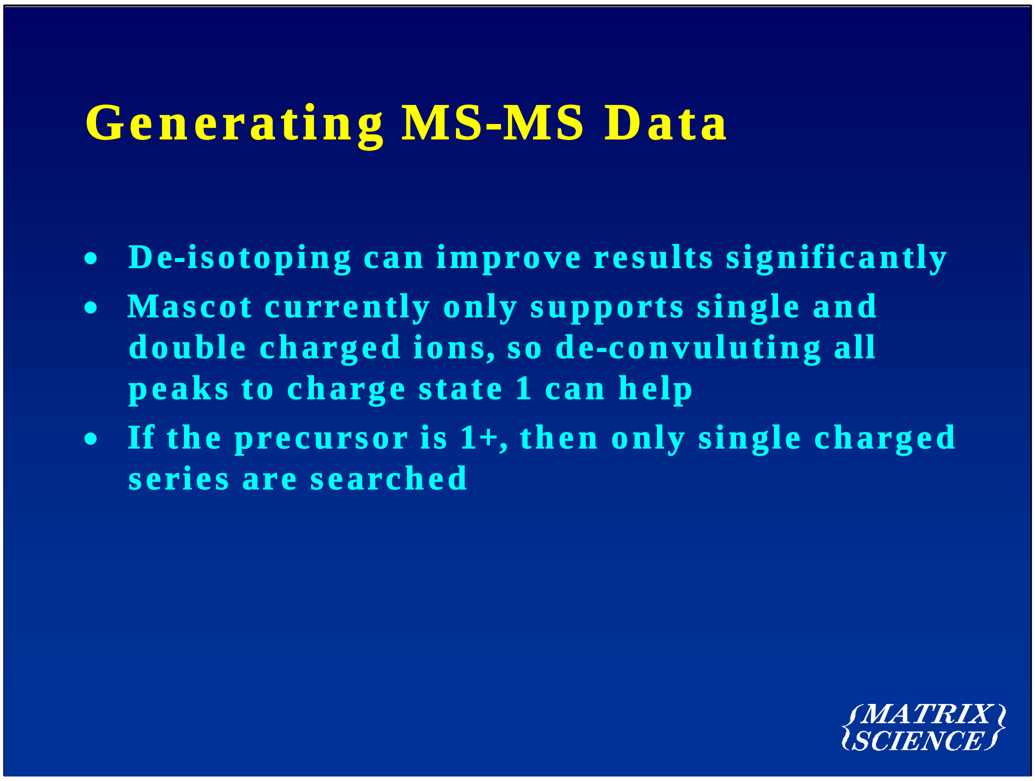
- Intensity values required and used
- Mascot limit of 10,000 peaks per spectrum

Ions Score: 48 **Matches (Bold Red):** 11/99 fragment ions using 21 most intense peaks

With MS-MS data, unlike peptide mass fingerprint data, the intensity values are required and used.

There is a limit of 10,000 peaks per spectrum - if you get a message that you have exceeded this limit, then it is likely that you are submitting profile data rather than centroided data.

Mascot is able to get reasonable matches from quite noisy data, as can be seen here, which is for the peak list - this is not profile data.



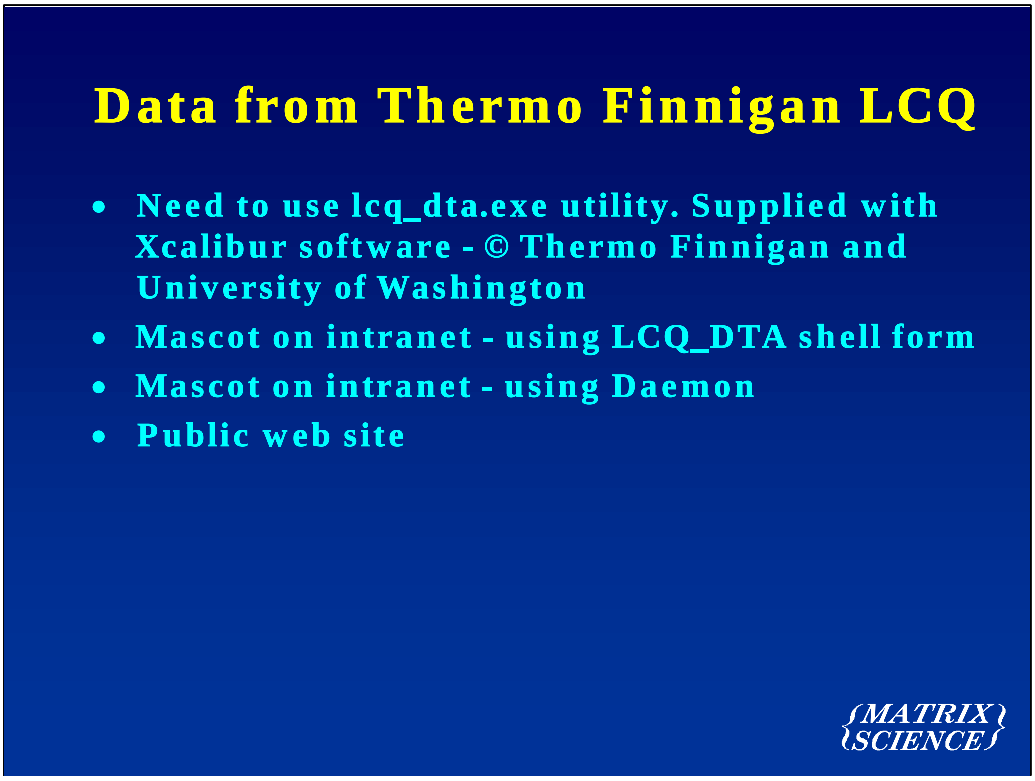
Generating MS-MS Data

- De-isotoping can improve results significantly
- Mascot currently only supports single and double charged ions, so de-convuluting all peaks to charge state 1 can help
- If the precursor is 1+, then only single charged series are searched

As with peptide mass fingerprint data, de-isotoping can improve results dramatically.

Mascot currently only supports singly and doubly charged ions, so de-convoluting to a single charge state will help when there are fragment ions with charge state 3 and higher.

If the precursor is singly charged, then only singly charged series are searched.



Data from Thermo Finnigan LCQ

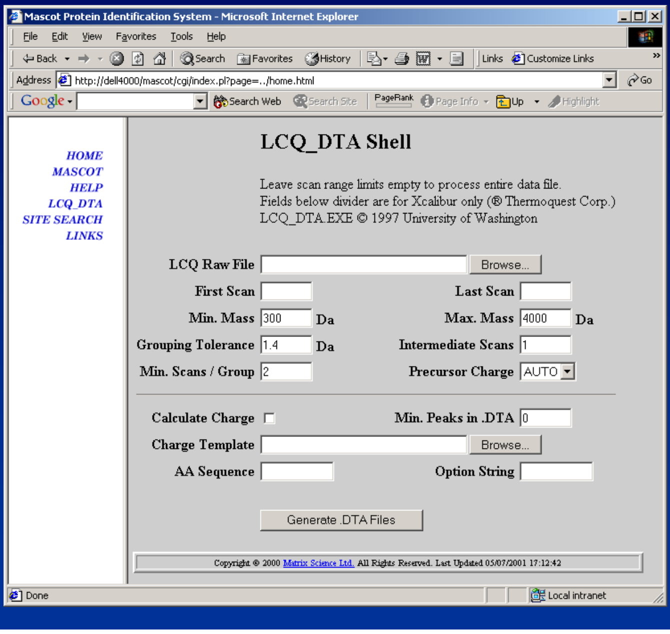
- Need to use lcq_dta.exe utility. Supplied with Xcalibur software - © Thermo Finnigan and University of Washington
- Mascot on intranet - using LCQ_DTA shell form
- Mascot on intranet - using Daemon
- Public web site

{MATRIX} {SCIENCE}

Some quick details now for specific instruments.

When using Mascot in-house, there are two tools to facilitate the transfer of data from ThermoFinnigan LCQ raw files into Mascot. These tools use the lcq_dta.exe utility provided with the Xcalibur - the LCQ_DTA browser form or Mascot Daemon.

LCQ_DTA Shell form


Mascot Protein Identification System - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites History Links Customize Links
Address http://dell4000/mascot/cgi/index.pl?page=../home.html
Go
Google Search Web Search Site PageRank Page Info Up Highlight
HOME
MASCOT
HELP
LCQ_DTA
SITE SEARCH
LINKS
Browse...
Browse...
Generate .DTA Files
Done
Local intranet

LCQ_DTA Shell

Leave scan range limits empty to process entire data file.
Fields below divider are for Xcalibur only (® Thermoquest Corp.)
LCQ_DTA.EXE © 1997 University of Washington

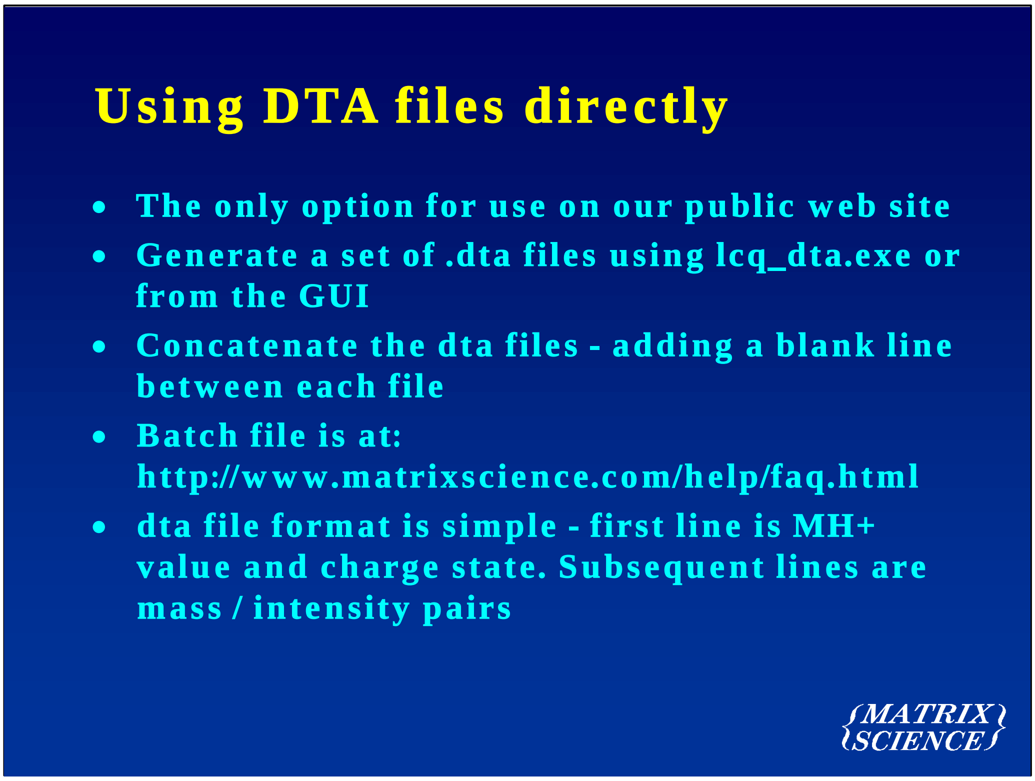
LCQ Raw File			
First Scan		Last Scan	
Min. Mass	300 Da	Max. Mass	4000 Da
Grouping Tolerance	1.4 Da	Intermediate Scans	1
Min. Scans / Group	2	Precursor Charge	AUTO
Calculate Charge	□	Min. Peaks in .DTA	0
Charge Template			
AA Sequence		Option String	

Copyright © 2000 Matrix Science Ltd. All Rights Reserved. Last Updated 05/07/2001 17:12:42

{MATRIX}
{SCIENCE}

If you click on the LCQ_DTA link on browser menu, then this screen will be displayed.

Choose your raw data file here. All the other parameters here are passed straight on to LCQ_DTA.EXE - you will need to refer to your LCQ documentation for details of these parameters.

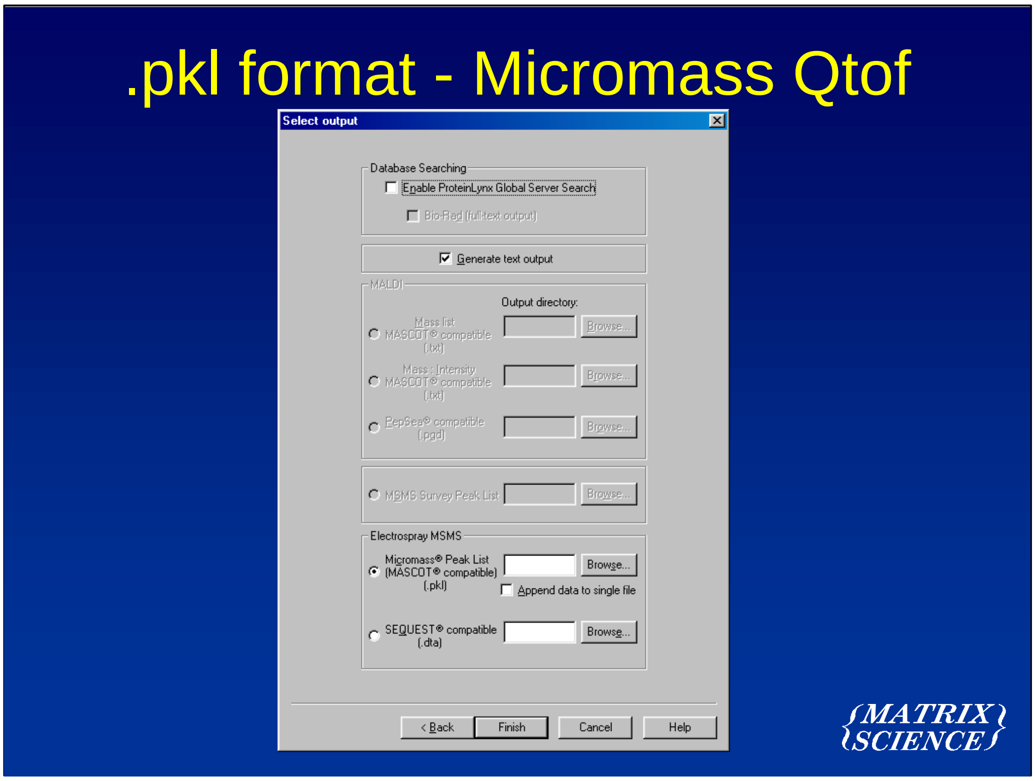


Using DTA files directly

- The only option for use on our public web site
- Generate a set of .dta files using lcq_dta.exe or from the GUI
- Concatenate the dta files - adding a blank line between each file
- Batch file is at: http://www.matrixscience.com/help/faq.html
- dta file format is simple - first line is MH+ value and charge state. Subsequent lines are mass / intensity pairs

These tools are not available on the web site. Here, you must create a set of .dta files using the Xcalibur GUI, or by using lcq_dta.exe directly. You then need to concatenate all the files, adding a blank line between them. You can either do this with an editor, or there is a batch file available to help.

The .dta file format is very simple - the first line is the precursor mass and charge state, and subsequent lines are the fragment mass and intensity pairs.



.pkl format - Micromass Qtof

Select output

Database Searching

☐ Enable ProteinLynx Global Server Search

☐ Bio-Rad (full-text output)

☑ Generate text output

MALDI

Output directory:

○ Mass list MASCOT® compatible (.txt) Browse...

○ Mass : Intensity MASCOT® compatible (.txt) Browse...

○ PepSea® compatible (.pgd) Browse...

○ MSMS Survey Peak List Browse...

Electrospray MSMS

● Micromass® Peak List (MASCOT® compatible) (.pkl) Browse...

☐ Append data to single file

○ SEQUEST® compatible (.dta) Browse...

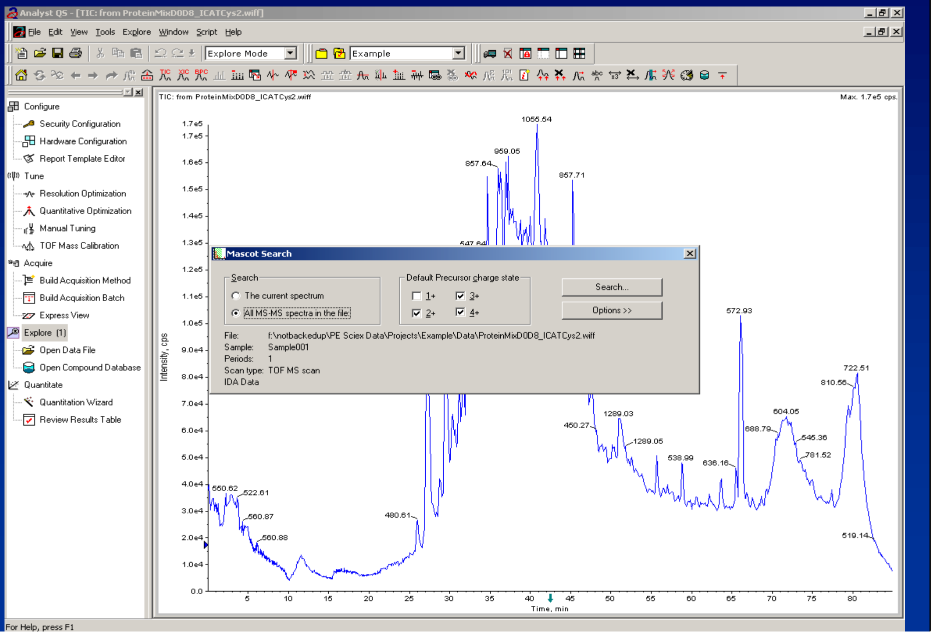
< Back | Finish | Cancel | Help

Masslynx .pkl files can be read by Mascot. They are almost identical to concatenated .dta files - the only difference is that the line with the precursor mass also includes the intensity.

You can export to .pkl files from ProteinLynx as shown in this dialog box. It is also possible to export Maldi lists from the linear M@ldi instrument using the same dialog box.

Applied Biosystems | MDS Sciex

Analyst software macro



Mascot Search

Search
○ The current spectrum
● All MS-MS spectra in the file

Default Precursor charge state
□ 1+ ☑ 3+
☑ 2+ ☑ 4+

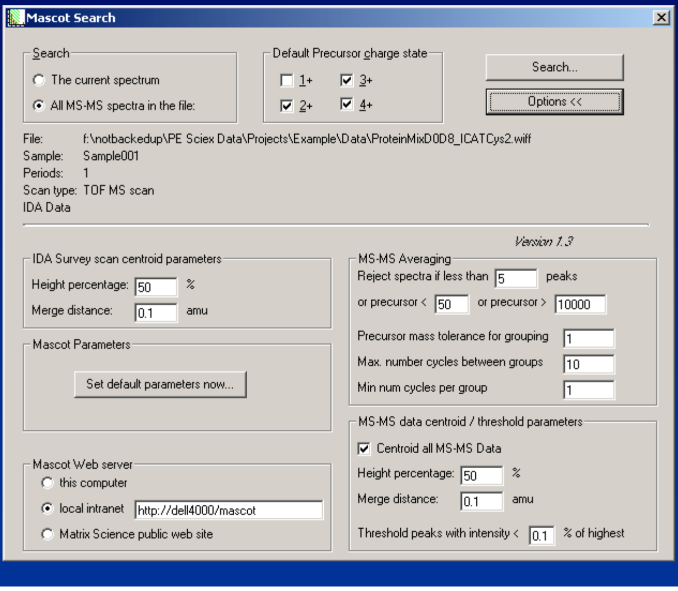
Search...
Options >>

File: f:\notbackedup\PE Sciex Data\Projects\Example\Data\ProteinMixD0D8_ICATCys2.wiff
Sample: Sample001
Periods: 1
Scan type: TOF MS scan
IDA Data

With Applied Biosystems MDS Sciex, there is a Mascot script that can be called from the Script menu. The macro is available from your local Sciex representative or from us.

Applied Biosystems | MDS Sciex

Analyst software macro - options



Mascot Search

Search
- ○ The current spectrum
- ● All MS-MS spectra in the file:

Default Precursor charge state
- □ 1+
- ☑ 2+
- ☑ 3+
- ☑ 4+

File: f:\notbackedup\PE Sciex Data\Projects\Example\Data\ProteinMixD0D8_ICATCys2.wiff
Sample: Sample001
Periods: 1
Scan type: TOF MS scan
IDA Data

Version 1.3

IDA Survey scan centroid parameters
- Height percentage: 50 %
- Merge distance: 0.1 amu

Mascot Parameters

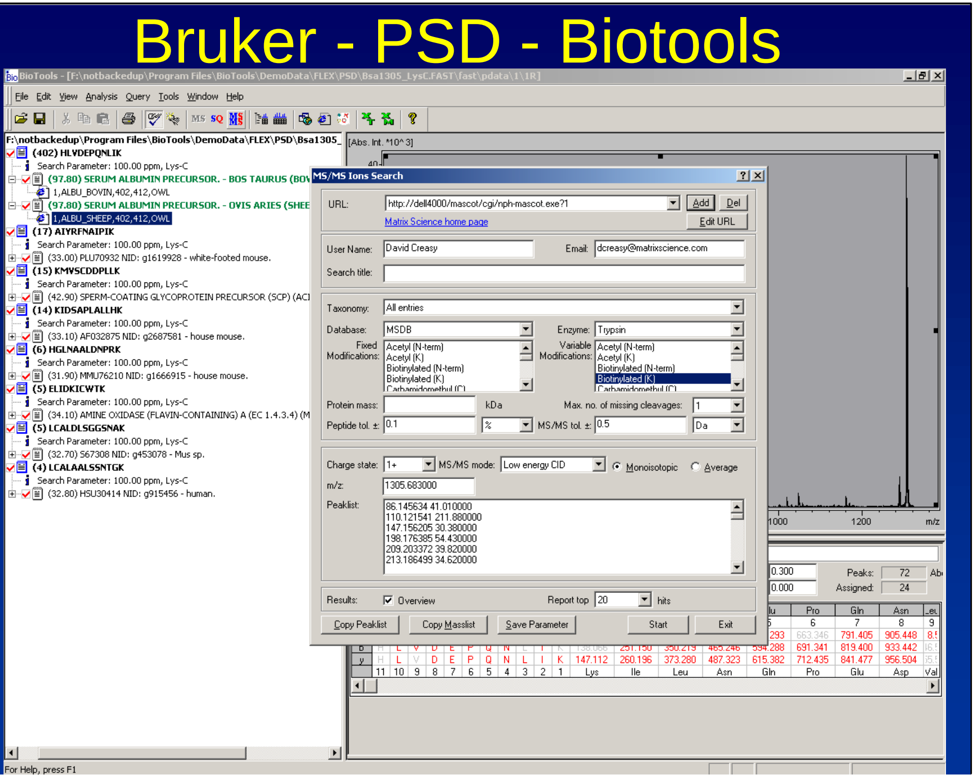
Mascot Web server
- ○ this computer
- ● local intranet http://dell4000/mascot
- ○ Matrix Science public web site

MS-MS Averaging
- Reject spectra if less than 5 peaks
- or precursor < 50 or precursor > 10000
- Precursor mass tolerance for grouping 1
- Max. number cycles between groups 10
- Min num cycles per group 1

MS-MS data centroid / threshold parameters
- ☑ Centroid all MS-MS Data
- Height percentage: 50 %
- Merge distance: 0.1 amu
- Threshold peaks with intensity < 0.1 % of highest

The options screen enables you to choose which Mascot server to use, and lots of the parameters that have just been discussed.

Bruker - PSD - Biotools



BioTools - [F:\notbackedup\Program Files\BioTools\DemoData\FLEX\PSD\Bsa1305_LysC.FAST\fast\pdata\1\1R]

F:\notbackedup\Program Files\BioTools\DemoData\FLEX\PSD\Bsa1305_

- (402) HLVDEPQNLIK
  - Search Parameter: 100.00 ppm, Lys-C
  - (97.80) SERUM ALBUMIN PRECURSOR. - BOS TAURUS (BOV
    - 1,ALBU_BOVIN,402,412,OWL
  - (97.80) SERUM ALBUMIN PRECURSOR. - OVIS ARIES (SHEE
    - 1,ALBU_SHEEP,402,412,OWL
- (17) AIYRFNAIPIK
  - Search Parameter: 100.00 ppm, Lys-C
  - (33.00) PLU70932 NID: g1619928 - white-footed mouse.
- (15) KMVSCDDPLLK
  - Search Parameter: 100.00 ppm, Lys-C
  - (42.90) SPERM-COATING GLYCOPROTEIN PRECURSOR (SCP) (ACI
- (14) KIDSAPLALLHK
  - Search Parameter: 100.00 ppm, Lys-C
  - (33.10) AF032875 NID: g2687581 - house mouse.
- (6) HGLNAALDNPRK
  - Search Parameter: 100.00 ppm, Lys-C
  - (31.90) MMU76210 NID: g1666915 - house mouse.
- (5) ELIDKICWTK
  - Search Parameter: 100.00 ppm, Lys-C
  - (34.10) AMINE OXIDASE (FLAVIN-CONTAINING) A (EC 1.4.3.4) (M
- (5) LCALDLSGGSNAK
  - Search Parameter: 100.00 ppm, Lys-C
  - (32.70) S67308 NID: g453078 - Mus sp.
- (4) LCALAALSSNTGK
  - Search Parameter: 100.00 ppm, Lys-C
  - (32.80) HSU30414 NID: g915456 - human.

MS/MS Ions Search

URL: http://dell4000/mascot/cgi/nph-mascot.exe?1

Matrix Science home page

User Name: David Creasy Email: dcreasy@matrixscience.com

Search title:

Taxonomy: All entries

Database: MSDB Enzyme: Trypsin

Fixed Modifications: Acetyl (N-term), Acetyl (K), Biotinylated (N-term), Biotinylated (K), Carbamidomethyl (C)

Variable Modifications: Acetyl (N-term), Acetyl (K), Biotinylated (N-term), Biotinylated (K), Carbamidomethyl (C)

Protein mass: kDa Max. no. of missing cleavages: 1

Peptide tol. ±: 0.1 % MS/MS tol. ±: 0.5 Da

Charge state: 1+ MS/MS mode: Low energy CID ● Monoisotopic ○ Average

m/z: 1305.683000

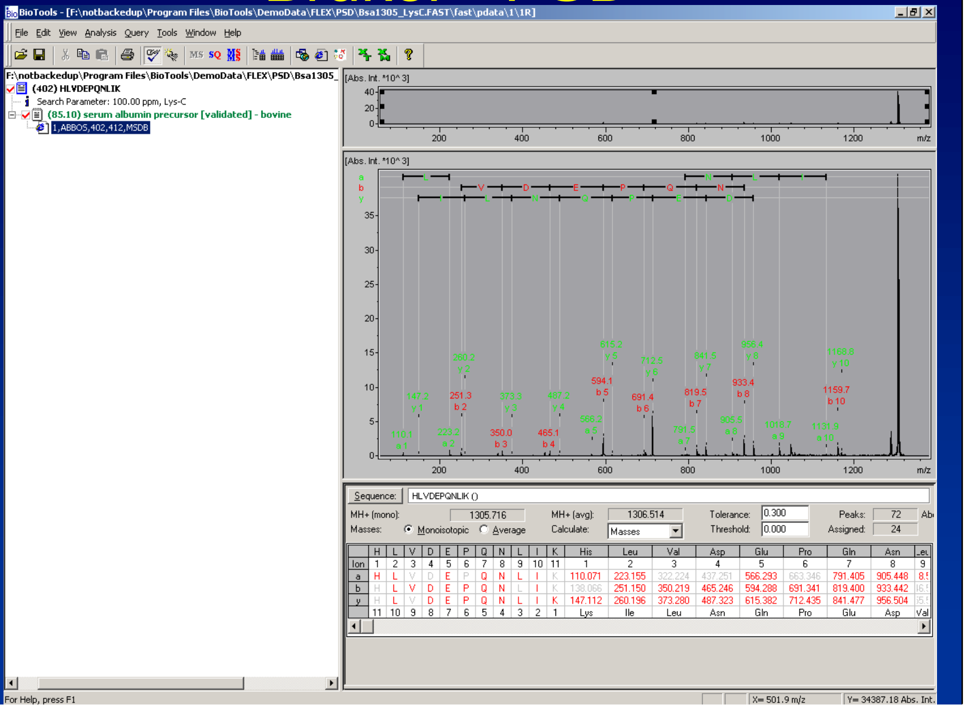
Peaklist:

```
86.145634 41.010000
110.121541 211.880000
147.156205 30.380000
198.176385 54.430000
209.203372 39.820000
213.186499 34.620000
```

Results: ☑ Overview Report top 20 hits

With BioTools from Bruker, there is very close integration with Mascot. A search form similar to the standard Mascot search form is available from the tool bar

Bruker - PSD



BioTools - [F:\notbackedup\Program Files\BioTools\DemoData\FLEX\PSD\Bsa1305_LysC.FAST\fast\pdata\1\1R]

F:\notbackedup\Program Files\BioTools\DemoData\FLEX\PSD\Bsa1305_

(402) HLVDEPQNLIK

Search Parameter: 100.00 ppm, Lys-C

(85.10) serum albumin precursor [validated] - bovine

1,ABBOS,402,412,MSDB

Sequence: HLVDEPQNLIK ()

MH+ (mono): 1305.716 MH+ (avg): 1306.514 Tolerance: 0.300 Peaks: 72 Ab

Masses: ● Monoisotopic ○ Average Calculate: Masses Threshold: 0.000 Assigned: 24

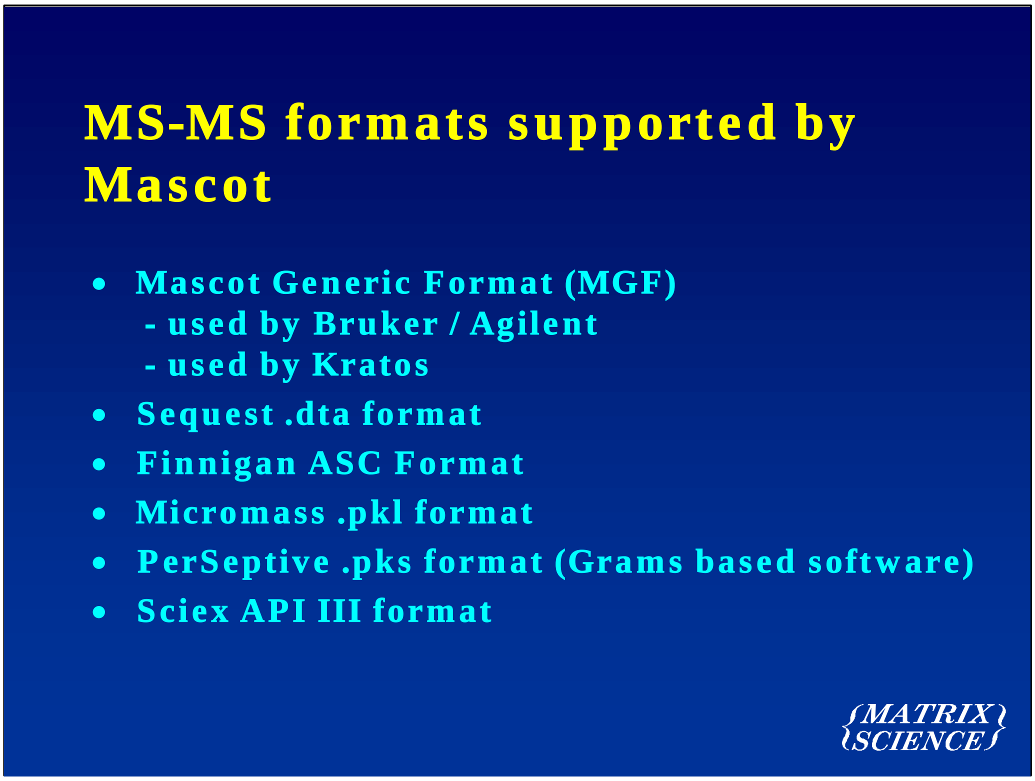
	H	L	V	D	E	P	Q	N	L	I	K	His	Leu	Val	Asp	Glu	Pro	Gln	Asn	Leu
Ion	1	2	3	4	5	6	7	8	9	10	11	1	2	3	4	5	6	7	8	9
a	H	L	V	D	E	P	Q	N	L	I	K	110.071	223.155	322.224	437.251	566.293	663.346	791.405	905.448	8.5
b	H	L	V	D	E	P	Q	N	L	I	K	138.066	251.150	350.219	465.246	594.288	691.341	819.400	933.442	16.5
y	H	L	V	D	E	P	Q	N	L	I	K	147.112	260.196	373.280	487.323	615.382	712.435	841.477	956.504	55.5
	11	10	9	8	7	6	5	4	3	2	1	Lys	Ile	Leu	Asn	Gln	Pro	Glu	Asp	Val

For Help, press F1

X= 501.9 m/z

Y= 34387.18 Abs. Int.

and the results are extracted from Mascot to be displayed back inside Biotools.



MS-MS formats supported by Mascot

- Mascot Generic Format (MGF)
  - used by Bruker / Agilent
  - used by Kratos
- Sequest .dta format
- Finnigan ASC Format
- Micromass .pkl format
- PerSeptive .pks format (Grams based software)
- Sciex API III format

The following file formats are supported by Mascot.

MGF is documented on our web site, and is used by some of the instrument manufacturers.

The Sequest .DTA format has already been discussed.

The Finnigan ASC format, is for files created by the LIST command on the old ICIS data system.

We have also seen that the Micromass .PKL format is supported.

Finally, these two older formats are also supported.

{MATRIX
SCIENCE}

http://www.matrixscience.com