Top 10 Tips for Successful Searching

ASMS 2003

SMATRIX SCIENCES

I'd like to present our top 10 tips for successful searching with Mascot. Like any hit parade, we will, of course, count them off in reverse order

10. Don't specify a poorly represented taxonomy

 In most cases, if the correct protein is not in the database, you'd like to see the closest match ... whatever the species Pie Edit Format Help
All entries=1437245
Archaea (Archaeobacteria)=48249
Eukaryota (eucaryotes)=679189
Alveolata (alveolates)=20844
Plasmodium falciparum (malaria parasite)=8561
Other Alveolata=12290
Metazoa (animals)=462646
Caenorhabditis elegans=27922
prosophila (fruit files)=29828
Chordata (vertebrates and relatives)=347716
Iobe=fined fish and tetrapod clade=322723
Mammalia (mamals)=293341
Primates=0900
Rodentia (Rodents)=145587
Mus.=96442
Mus musculus (house mouse)=95155
Rattus=43907
Other rodentia=5807
Other mammalia=27359

✓

ASMS 2003



So, at number 10, Don't specify a poorly represented taxonomy.

Think carefully about what you are trying to achieve when specifying a taxonomy filter.

If the correct protein from the correct species is not in the database, wouldn't you want to see a good match to a protein from a different species?

This is especially important for poorly represented species. For example, look at these numbers for the NCBI nr database in June 2003: 1.4 million entries; 120,000 entries for primates, of which all but 9,000 are for human. So, even if you are studying chimps or orang-utans or yeti, you probably don't want to choose 'Other primates'.

9. Use the Peptide Summary Report for MS/MS results

- The Protein Summary Report is intended for Peptide Mass Fingerprint results
- Worst case is a complex mixture with lots of queries
 - Protein Summary is 50 proteins max
 - Matches to sets of identical peptides are not collapsed into single protein hits, so a match may disappear off the end of the top 50
 - Weak matches may disappear into the distribution of random PMF matches

ASMS 2003



At number 9, We encourage you to Use the Peptide Summary Report for MS/MS results

There are several flavours of reports for Mascot search results. Historically, the first report was the Protein Summary, used for peptide mass fingerprint results. Because this was the first report, there are still some old clients out there that specify this report for all searches. Unfortunately, in most cases, the Protein Summary is not a good way to view MS/MS results. For example:

| Mascot Search Results - Microsoft Internet Ex | | | _ 8 3 |
|--|--|------------|-----------|
| File Edit View Favorites Tools Help | μοι τ | | |
| | | | |
| | vorites 🛞 Media 🥨 🛃 🖆 🖾 🗐 🕅 | | N |
| Address 🕘 http://www.matrixscience.com/cgi/master_ | results.pl?file=/data/20030531/FTncixYE.dat | • | ∂Go Links |
| (MATRIX) SCIENCE/ Mascot Sear | ch Results | | |
| Vser : Email : Search title : | | | |
| MS data file : | | | |
| | (1434562 sequences; 461081222 residues) | | |
| Timestamp : 31 May 2003 at : Significant hits: gi 13623199 | L4:59:30 GMT ATP citrate lyase [Homo sapiens] | | |
| gi 418694 | serum albumin precursor [validated] - bovine | | |
| gi 11544639 | importin7 [Homo sapiens] | | |
| gi 2102696 | karyopherin beta 3 [Homo sapiens] | | |
| <u>gi 1351908</u> gi 6136092 | Serum albumin precursor (Allergen Fel d 2) Ubiquitin-activating enzyme E1 | | |
| gi 1362855 | lrp protein - human | | |
| gi 2244652 | HS24/P52 [Homo sapiens] | | |
| gi 3287937 | Vinculin (Metavinculin) | | |
| <u>gi 4501841</u> gi 1082769 | alanyl-tRNA synthetase [Homo sapiens] RNA helicase A - human | | |
| gi 4240147 | KIAA0829 protein [Homo sapiens] | | |
| gi 191765 | alpha-fetoprotein | | |
| <u>gi 28194088</u> | karyopherin-beta 3 variant [Xenopus laevis] | | |
| Probability Based Mowse Score | | | |
| Score is -10*Log(P), where P is the probab Individual ions scores > 48 indicate identity | ility that the observed match is a random event. | | |
| monorous fons scores > 40 monoate identity (| n extensive noniorogy (p∼0.05). | | |
| 11 the contract of the contrac | | | |
| 10 | | | |
| | | | Þ |
| | | 🔮 Internet | |

This result from an MS/MS search has 12 significant matches. There is a little bit of duplication, e.g. 2 serum albumins, but not much

| _ | scot Search Results - M | | xplorer | -8- |
|----------|--|----------------------------|-----------------|--|
| Eile | <u>E</u> dit <u>V</u> iew F <u>a</u> vorites | <u>T</u> ools <u>H</u> elp | | |
| 🕁 Bai | :k • 🔿 - 🙆 🛃 🚮 | 🛛 🔕 Search 🛛 🝙 F | avorites 🍕 | Media 🧭 🛃 - 🎒 🗹 📃 🕼 |
| ddres | s 🗿 http://www.matrixse | rience.com/cgi/master | r results.nl?fi | e=/data/20030531/FTncixYE.dat&REPTYPE=protein&REPORT=50 |
| <u>_</u> | | cloricomy egymascor | _rosacsipiti | |
| | | | | |
| Ind | ex | | | |
| | | | _ | |
| | Accession | Mass | Score | Description |
| | gi 20141248 | 120748 | 368 363 | ATP-citrate synthase (ATP-citrate (pro-S-)-lyase) (Citrate cleavage enzyme) |
| | gi 13623199 gi 8392839 | 120762 120559 | 363 292 | ATP citrate lyase [Homo sapiens] ATP citrate lyase [Rattus norvegicus] |
| | | 119651 | 292 | |
| | gi 28514402 gi 4501865 | 121342 | 292 | ATP citrate lyase [Mus musculus] ATP citrate lyase [Homo sapiens] |
| | gi 17028103 | 92420 | 223 | ATP-citrate lyase [Rattus norvegicus] |
| | gi 18204829 | 64981 | 223 | Acly protein [Mus musculus] |
| | gi 21754275 | 76396 | 213 | unnamed protein product [Homo sapiens] |
| | gi 5851949 | 18527 | 189 | ATP-citrate lyase [Gallus gallus] |
| | gi 2190337 | 69278 | 153 | serum albumin [Bos taurus] |
| | gi 418694 | 69225 | 153 | serum albumin precursor [validated] - bovine |
| | gi 1351907 | 69248 | 153 | Serum albumin precursor (Allergen Bos d 6) |
| | gi 27679544 | 98507 | 131 | similar to RAN binding protein 7; RAN binding protein 7 (importin 7); importin 7 [|
| | gi 26333317 | 103295 | 129 | unnamed protein product [Mus musculus] |
| | gi 11544639 | 116304 | 129 | importin7 [Homo sapiens] |
| | gi 1351908 | 68615 | 123 | Serum albumin precursor (Allergen Fel d 2) |
| | gi 5453998 | 119440 | 127 | importin 7; RAN-binding protein 7 [Homo sapiens] |
| | gi 11342591 | 119509 | 127 | RanBP7/importin 7 [Mus musculus] |
| | gi 28481575 | 82681 | 126 | similar to RAN binding protein 7; RAN binding protein 7 (importin 7); importin 7 [|
| | gi 2102696 | 123512 | 120 | karyopherin beta 3 [Homo sapiens] |
| | gi 28277071 | 125507 | 120 | karyopherin (importin) beta 3 [Homo sapiens] |
| | gi 24797086 | 125464 | 120 | karyopherin beta 3; Ran GTP binding protein 5; importin beta-3 subunit [Homo sapie |
| | gi 4033763 | 123550 | 120 | Importin beta-3 subunit (Karyopherin beta-3 subunit) (Ran-binding protein 5) |
| | gi 416704 | 186013 | 113 | Balbiani RING protein 3 precursor |
| | gi 3319897 | 65861 | 104 | albumin [Canis familiaris] |
| | gi 6687188 | 68560 | 101 | albumin [Canis familiaris] |
| | gi 2147092 | 29989 | 94 | albumin - dog (fragment) |
| | gi 229552 | 66088 | 94 | albumin |
| | gi 13124699 | 68562 | 90 | Serum albumin precursor (Allergen Can f 3) |
| | qi 1314732 | 186032 | 87 | 185 kDa silk protein |
| | gi 20826641 | 106352 | 87 | similar to zinc finger protein 91 (HPF7, HTF10) [Homo sapiens] [Mus musculus] |
| | gi 2244652 | 52334 | 83 | H324/P52 [Homo saniens] |
| | gi 164318 | 69352 | 81 | albumin |
| | gi 113578 | 69366 | 81 | Serum albumin precursor |
| | gi 13278232 | 73717 | 81 | heat shock protein, 110 kDa [Mus musculus] |
| | | | | |
| | ie | | | |

If we look at the same results in protein view, there is much greater duplication, because this type of report isn't trying to collapse hits that share a common set of MS/MS matches.

Now, we have 7 or 8 representatives for the more common protein hits, which means that the lower scoring hits are pushed off the bottom of the list.

Also, you can't see the wood for the trees.

So, if you have old client software that brings up a protein summary for an MS/MS search, the first thing to do is click on the link to switch to the peptide summary

8. Submit new modifications to Unimod

On-line at www.unimod.org

- Saves calculating mass values
- Saves having to understand the syntax of the Mascot mod_file
- Share your modification with other Mascot users
- Provides a way to update the modifications list on the Mascot public web site

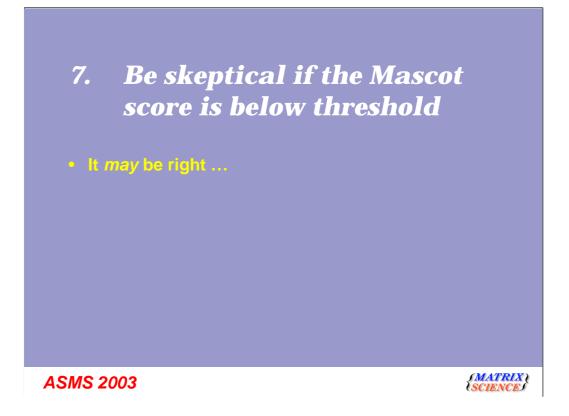
ASMS 2003



At number 8, if the modification you need isn't on the Mascot search form list, submit it to Unimod. The advantages of doing this are <read from slide>

| Edit ⊻iew Fa | | | | | |
|-----------------------------|-----------------------|--|---------------------------------|--------------------------------|-----------------------------------|
| ack 🔹 🔿 👻 🙆 | 😰 🚮 🥘 Search | 📓 Favorites 🛛 🖓 History 🛛 🖓 🕶 🎒 | Powermarks 🕅 🗡 | \$ [*] /• | |
| ss 🩋 columns_to_ | _view=code_name&colur | nns_to_view=mono_mass&columns_to_vie | ew=avge_mass&columns_to_view=co | mposition&display_details_view | v.x=7&display_details_view.y=13 🗾 |
| | | | | | |
| UNIA | | in modifications for mass spe | ctrometry | | |
| View All Records | | | | Help Home Opt | ions Advanced Search Logon |
| | 100 1100010 | | | | Gn |
| | | | | Search | |
| | | | | | |
| Record Detai | | | And the last | | |
| Accession # | | Short name | Acetyl_light | | |
| Modification Composition | | reagent light form (N-term 8 Monoisotopic | 42.010565 | Average | 42.0367 |
| Specificity De | | monoisocopic | 42.010505 | Average | 42.0307 |
| | K | Position | Anywhere | | |
| Neutral Loss | | Monoisotopic | , | Average | |
| | Isotopic label | | | | |
| Comment | | | | | |
| Specificity De | efinition 2 | | | | |
| Site | N-term | Position | Any N-term | | |
| Neutral Loss | | Monoisotopic | | Average | |
| | Isotopic label | | | | |
| Comment Notes and Re | afarancas | | | | |
| Source | | ce Controlling Deuterium isot | tone effects in comparativ | e proteomics. Zhang | Routian: Sioma Cathy |
| o di Go | Joannan Const en | S.; Thompson, Robert A.; University, West Lafayette | ; Xiong, Li; Regnier, Fred I | E. Department of Ch | |
| Source | Journal Referen | ce Global internal standard to Fred E Department of Ch Chromatography, A (2002 | nemistry, Purdue Universit | | |
| Source | Journal Referen | ce Comparative proteomics E Riggs, Larry; Zhang, Rou Cathy; Thompson, Robert | jian; Xiong, Li; Liu, Peiran; | : Chakraborty, Asish; | |
| Notes | | | | | |
| Curator | penner | Last Modified | 2002-10-20 10:5 | 0:36 | |
| | | | | | Email Change Request |

Unimod is a live, public domain database. If you add a modification, you become the curator of that modification. The database is used to update the Mascot mod_file every weekend. If you have an in-house Mascot server, you can download the same new mod_file



At number 7, with a bullet, Be skeptical if you want to accept a match when the Mascot score is below threshold. You may be right ...

| | <u>E</u> dit <u>V</u> i | iew F <u>a</u> vol | | ols <u>H</u> elp | | | | | | | | | | |
|-----------------------------------|--|--|---|---------------------|---|---|---|--|---|---|--|--|-----------------------|---|
| | Back → ⇒ | | | 🗟 Search | _ | | | | | 1 | marks 🕅 | | • | |
| Įdre | ess 🤌 http | p://www.ma | atrixscience | e.com/cgi/pe | ptide_view. | pl?file=/o | data/200008 | 28/FTm | icobu.dat&q | uery=338ł | nit=38index: | =8p× | =1 | - |
| | | | | | y(8) | | | | | | | | | |
| | | | | | 62 | | | | | | | | | |
| | | | | | | | | | | | | | | _ |
| | | 6 | y(4) | ~ | | -1010) - y(9) | | | | | | (| | |
| | | | | (9)h- | | a | | | | | | | Identity threshold | |
| | | | | | | 1) di | _ | | | | | | - | |
| | b(2)++,L | a(4) | 9 | | | | 9(11) A | | | | | | for this search is 73 | |
| | :)q | | .(9)#d- | | | | 1 | | | | | | | |
| ŗ | يها للنسب | ملوا فاستكالك | الموا الحق | فتحجم فالقبر للسل | htere and the second second | عللم بلاب | | | ····· | <u> </u> | | | | |
| 0 | | 400 | | 800 | | 1200 | | 1600 | | 2000 | | | | |
| | | | | utral pe (Bold P | | | 426.86 agment : | ions | using 3 | 9 most | intense | e pe | eaks | |
| DN | s Score | : 55 M | atches | (Bold P | ted): 19 | /89 fr | agment : | | - | | | _ | eaks | |
| on # | s Score Immon. | :55 M | atches | | b | /89 fr b ⁺⁺ | agment : | Seq. | using 3 | 9 most y ⁺⁺ | intense y* | # | zaks | |
| 9n ¥ 1 | s Score Immon. 86.10 | : 55 M a <u>86.10</u> | atches a⁺⁺ 43.55 | (Bold P | b 114.09 | /89 fr b ⁺⁺ 57.55 | agment : | Seq. L | У | y ⁺⁺ | y* | # 12 | zaks | |
| סח 4 1 2 | s Score Immon. 86.10 30.03 | a 86.10 143.12 | atches a ⁺⁺ 43.55 72.06 | (Bold P | b 114.09 171.11 | /89 fr b ⁺⁺ 57.55 86.06 | agment : | Seq. L G | y 1314.79 | y ++ 657.90 | y* 1297.76 | # 12 11 | zaks | |
| 9 n # 1 2 3 | s Score Immon. 86.10 30.03 74.06 | a 86.10 143.12 244.17 | atches a ⁺⁺ 43.55 72.06 122.59 | (Bold P | b 114.09 171.11 272.16 | b ⁺⁺ 57.55 86.06 136.58 | agment : | Seq. L G T | y 1314.79 1257.76 | y ⁺⁺ 657.90 629.39 | y* 1297.76 1240.74 | # 12 11 10 | zaks | |
|) 1 2 3 4 | s Score Immon. 86.10 30.03 74.06 86.10 | a 86.10 143.12 244.17 357.25 | a ⁺⁺ 43.55 72.06 122.59 179.13 | (Bold P | b 114.09 171.11 272.16 385.25 | /89 fr b ⁺⁺ 57.55 86.06 136.58 193.13 | agment : | Seq. L G T L | y 1314.79 1257.76 1156.72 | y ⁺⁺ 657.90 629.39 578.86 | y* 1297.76 1240.74 1139.69 | # 12 11 10 9 | zaks | |
| 9 n # 1 2 3 4 5 | s Score Immon. 86.10 30.03 74.06 86.10 102.06 | e 55 M a 86.10 143.12 244.17 357.25 486.29 | a ⁺⁺ 43.55 72.06 122.59 179.13 243.65 | (Bold F | b 114.09 171.11 272.16 385.25 514.29 | /89 fr b ⁺⁺ 57.55 86.06 136.58 193.13 257.65 | b* | Seq. L G T L E | y 1314.79 1257.76 1156.72 1043.63 | y ⁺⁺ 657.90 629.39 578.86 522.32 | y* 1297.76 1240.74 1139.69 1026.61 | # 12 11 10 | zaks | |
| on | s Score Immon. 86.10 30.03 74.06 86.10 102.06 101.11 | : 55 M 86.10 143.12 244.17 357.25 486.29 614.39 | atches a ⁺⁺ 43.55 72.06 122.59 179.13 243.65 307.70 | (Bold F | b 114.09 171.11 272.16 385.25 514.29 642.38 | <pre>/89 fr b++ 57.55 86.06 136.58 193.13 257.65 321.70</pre> | b* | Seq. L G T L E K | y 1314.79 1257.76 1156.72 1043.63 914.59 | y ⁺⁺ 657.90 629.39 578.86 522.32 457.80 | y* 1297.76 1240.74 1139.69 1026.61 897.56 | # 12 11 10 9 8 7 | zaks | |
| 2 3 4 5 6 7 | s Score Immon. 86.10 30.03 74.06 86.10 102.06 101.11 87.06 | : 55 M 86.10 143.12 244.17 357.25 486.29 614.39 728.43 | atches a ⁺⁺ 43.55 72.06 122.59 179.13 243.65 307.70 364.72 | (Bold F a* | b 114.09 171.11 272.16 385.25 514.29 642.38 756.43 | <pre>/89 fr b⁺⁺ 57.55 86.06 136.58 193.13 257.65 321.70 378.72</pre> | agment : b* 625.36 739.40 | Seq. L G L L E K N | y 1314.79 1257.76 1156.72 1043.63 914.59 786.49 | y ⁺⁺ 657.90 629.39 578.86 522.32 457.80 393.75 | y* 1297.76 1240.74 1139.69 1026.61 897.56 769.47 | # 12 11 10 9 8 7 6 | taks | |
| <pre>m # 1 2 3 4 5 6 7 8</pre> | s Score Immon. 86.10 30.03 74.06 86.10 102.06 101.11 87.06 101.07 | : 55 M | atches a ⁺⁺ 43.55 72.06 122.59 179.13 243.65 307.70 364.72 428.75 | (Bold F a* | b 114.09 171.11 272.16 385.25 514.29 642.38 756.43 884.48 | b++ 57.55 86.06 136.58 193.13 257.65 321.70 378.72 442.75 | agment : b* 625.36 739.40 867.46 | Seq. L G T L E K N | y 1314.79 1257.76 1156.72 1043.63 914.59 786.49 672.45 | y ⁺⁺ 657.90 629.39 578.86 522.32 457.80 393.75 336.73 | y* 1297.76 1240.74 1139.69 1026.61 897.56 769.47 655.43 | # 12 11 10 9 8 7 6 5 | zaks | |
| <pre>pn # 1 2 3 4 5 6 7 8 9</pre> | S Score Immon. 86.10 30.03 74.06 86.10 102.06 101.11 87.06 101.07 101.11 | : 55 M 86.10 143.12 244.17 357.25 486.29 614.39 728.43 856.49 984.58 | atches a ⁺⁺ 43.55 72.06 122.59 179.13 243.65 307.70 364.72 428.75 492.80 | (Bold F a* | b 114.09 171.11 272.16 385.25 514.29 642.38 756.43 884.48 1012.58 | b++ 57.55 86.06 136.58 193.13 257.65 321.70 378.72 442.75 506.79 | agment : b* 625.36 739.40 867.46 995.55 | Seq. G T L E K Q K | y 1314.79 1257.76 1156.72 1043.63 914.59 786.49 672.45 544.39 | y ⁺⁺ 657.90 629.39 578.86 522.32 457.80 393.75 336.73 272.70 | y* 1297.76 1240.74 1139.69 1026.61 897.56 769.47 655.43 527.37 | # 12 11 10 9 8 7 6 5 4 | zaks | |
| n | S Score Immon. 86.10 30.03 74.06 86.10 102.06 101.11 87.06 101.07 101.11 101.11 | : 55 M 86.10 143.12 244.17 357.25 486.29 614.39 728.43 856.49 984.58 1112.68 | atches a ⁺⁺ 43.55 72.06 122.59 179.13 243.65 307.70 364.72 428.75 492.80 556.84 | (Bold F a* | b 114.09 171.11 272.16 385.25 514.29 642.38 756.43 884.48 1012.58 1140.67 | b++ 57.55 86.06 136.58 193.13 257.65 321.70 378.72 442.75 506.79 570.84 | agment : b* 625.36 739.40 867.46 995.55 1123.65 | Seq. L G T L E K N Q K K | y 1314.79 1257.76 1156.72 1043.63 914.59 786.49 672.45 544.39 416.30 | y ⁺⁺ 657.90 629.39 578.86 522.32 457.80 393.75 336.73 272.70 208.65 | y* 1297.76 1240.74 1139.69 1026.61 897.56 769.47 655.43 527.37 399.27 | # 12 11 10 9 8 7 6 5 4 3 | zaks | |
| <pre>m # 1 2 3 4 5 6 7 8</pre> | S Score Immon. 86.10 30.03 74.06 86.10 102.06 101.11 87.06 101.07 101.11 101.11 | : 55 M 86.10 143.12 244.17 357.25 486.29 614.39 728.43 856.49 984.58 1112.68 | atches a ⁺⁺ 43.55 72.06 122.59 179.13 243.65 307.70 364.72 428.75 492.80 556.84 | (Bold F a* | b 114.09 171.11 272.16 385.25 514.29 642.38 756.43 884.48 1012.58 1140.67 | b++ 57.55 86.06 136.58 193.13 257.65 321.70 378.72 442.75 506.79 570.84 | agment : b* 625.36 739.40 867.46 995.55 1123.65 | Seq. G T L E K Q K | y 1314.79 1257.76 1156.72 1043.63 914.59 786.49 672.45 544.39 416.30 | y ⁺⁺ 657.90 629.39 578.86 522.32 457.80 393.75 336.73 272.70 208.65 144.61 | y* 1297.76 1240.74 1139.69 1026.61 897.56 769.47 655.43 527.37 | # 12 11 10 9 8 7 6 5 4 3 | zaks | |

Here's a good example.

A run of 9 Y ions. Who wants to tell me that this could never happen by chance? And yet the score is below threshold!

OK, now lets take a look at a different match

| ⁼ile | | ch Results jew F <u>a</u> voi | | View - Mici | rosoft Inte | ernet Exp | lorer | | | | | | | | | |
|---|--|---|---|-----------------------------------|---|---|--------------------|--|---|---|--|---|----------------|--|---|------|
| | | | | Search | Gel Esucrito | | toru I EL | - 6 | | Bouror | marks 🎼 | A ' | × | | | |
| | | | | | | | | | | - | | | 7c1422899&px=1 | | • | - R |
| | se le na | p.//www.inc | scixscience | steomyegiype | | pinioye | 33(3)200000 | 2071 11 | icoba.aacoq | der y=3501 | IC-TOINGSX | -gi 70 | | | |] (r |
| | | | | | -9(8) | | | | | | | | | | | |
| | | | | | 9(7) | | | | | | | | | | | |
| | | | ~ | | i | â | | | | | | | | | | |
| | ~ | | -y(4) | (9)h | | | | | | | | | | | | |
| | - 40 | 8 | | 5 5 | | (01) | | | | | | | | | | |
| | a(2) - y(1) | a(4) | | | | - y(10) | y(11) | | | | | | | | | |
| | | a. | | | | , n., | 5 | | | | | | | | | |
| | | h h h h h h h h h h h h h h h h h h h | للسف الحس | المرجعة الأراقية | السالي | | | | | | | | | | | |
| 5 | | 400 | | 800 | | 1200 | | 1600 | | 2000 | | | | | | |
| | | | | | | | | | | | | | | | | |
| - | | | | | | | | | | | | | | | | |
| | | | | utral pe | | | | | | | | | | | | |
| | | | | utral pe (<mark>Bold F</mark> | | | | ions | using 3 | 9 most | intense | e pe | aks | | | |
| n | | | | | | | agment : | ions Seq. | using 3 | 9 most | intense y* | e pe # | aks | | | |
| n : | s Score | : 83 M | atches | (Bold P | led): 20 |)/87 fr b ⁺⁺ | agment : | | - | | | _ | aks | | | |
| n | s Score Immon. | a 86.10 | atches a ⁺⁺ 43.55 | (Bold P | b 114.09 |)/87 fr b ⁺⁺ | agment : | Seq. | У | y ⁺⁺ | | # 12 | aks | | | |
| n: [| s Score Immon. 86.10 | a 86.10 173.13 | atches a ⁺⁺ 43.55 87.07 | (Bold P | b 114.09 201.12 | b ⁺⁺ 57.55 | agment : b* | Seq. I | y 1314.70 | y ++ 657.85 | y* | # 12 11 | aks | | | |
| n: L 2 | s Score Immon. 86.10 60.04 44.05 | a 86.10 173.13 244.17 | atches a ⁺⁺ 43.55 87.07 122.59 | (Bold P | b 114.09 201.12 272.16 | b ⁺⁺ 57.55 101.07 | agment : b* | Seq. I S | y 1314.70 1227.67 | y ⁺⁺ 657.85 614.34 | y* 1297.68 | # 12 11 10 | aks | | | |
| n: L 2 3 | s Score Immon. 86.10 60.04 44.05 86.10 | a 86.10 173.13 244.17 | atches 43.55 87.07 122.59 179.13 | (Bold P | b 114.09 201.12 272.16 385.25 | b ⁺⁺ 57.55 101.07 136.58 | b* | Seq. I S A | y 1314.70 1227.67 1156.63 | y ⁺⁺ 657.85 614.34 578.82 | y* 1297.68 1210.64 | # 12 11 10 9 | aks | | | |
| n: L 2 5 | s Score Immon. 86.10 60.04 44.05 86.10 102.06 | a 86.10 173.13 244.17 357.25 | atches a ⁺⁺ 43.55 87.07 122.59 179.13 243.65 | (Bold F | b 114.09 201.12 272.16 385.25 514.29 | b ⁺⁺ 57.55 101.07 136.58 193.13 | b* | Seq. I S A L | y 1314.70 1227.67 1156.63 1043.55 | y ⁺⁺ 657.85 614.34 578.82 522.28 | y* 1297.68 1210.64 1139.61 | # 12 11 10 9 8 | aks | | | |
| n 1 2 3 4 5 | s Score Immon. 86.10 60.04 44.05 86.10 102.06 102.06 | a 86.10 173.13 244.17 357.25 486.29 615.34 | atches a ⁺⁺ 43.55 87.07 122.59 179.13 243.65 308.17 | (Bold F | b 114.09 201.12 272.16 385.25 514.29 | b ⁺⁺ 57.55 101.07 136.58 193.13 257.65 322.17 | b* | Seq. I S A L E E | y 1314.70 1227.67 1156.63 1043.55 914.51 | y ⁺⁺ 657.85 614.34 578.82 522.28 457.76 | y* 1297.68 1210.64 1139.61 1026.52 | # 12 11 10 9 8 7 | aks | | | |
| n L 2 3 4 5 5 7 | s Score Immon. 86.10 60.04 44.05 86.10 102.06 102.06 102.06 101.07 | a 86.10 173.13 244.17 357.25 486.29 615.34 743.39 | atches a ⁺⁺ 43.55 87.07 122.59 179.13 243.65 308.17 372.20 | (Bold F | b 114.09 201.12 272.16 385.25 514.29 643.33 771.39 | b/87 fr b ⁺⁺ 57.55 101.07 136.58 193.13 257.65 322.17 386.20 | agment : | Seq. I S A L E E Q | y 1314.70 1227.67 1156.63 1043.55 914.51 785.46 | y ⁺⁺ 657.85 614.34 578.82 522.28 457.76 | y* 1297.68 1210.64 1139.61 1026.52 897.48 768.44 | # 12 11 10 9 8 7 6 | aks | | | |
| n 2 3 4 5 7 3 | Score Immon. 86.10 60.04 44.05 86.10 102.06 102.06 101.07 86.10 | a 86.10 173.13 244.17 357.25 486.29 615.34 743.39 856.48 | atches a ⁺⁺ 43.55 87.07 122.59 179.13 243.65 308.17 372.20 428.74 | (Bold F a* | ked): 20 b 114.09 201.12 272.16 385.25 514.29 643.33 771.39 884.47 | b ⁺⁺ 57.55 101.07 136.58 193.13 257.65 322.17 386.20 442.74 | agment : b* | Seq. I S A L E Q L | y 1314.70 1227.67 1156.63 1043.55 914.51 785.46 657.40 | y ⁺⁺ 657.85 614.34 578.82 522.28 457.76 393.24 | y* 1297.68 1210.64 1139.61 1026.52 897.48 768.44 640.38 | # 12 11 10 9 8 7 6 5 | aks | | | |
| n L 2 3 4 5 5 7 3 | Score Immon. 86.10 60.04 44.05 86.10 102.06 102.06 101.07 86.10 101.07 | a 86.10 173.13 244.17 357.25 486.29 615.34 743.39 856.48 984.54 | atches a ⁺⁺ 43.55 87.07 122.59 179.13 243.65 308.17 372.20 428.74 492.77 | (Bold F a* | b 114.09 201.12 272.16 385.25 514.29 643.33 771.39 884.47 1012.53 | b++ 57.55 101.07 136.58 193.13 257.65 322.17 386.20 442.74 506.77 | agment : b* | Seq. I S A L E Q L Q | y 1314.70 1227.67 1156.63 1043.55 914.51 785.46 657.40 544.32 | y ⁺⁺ 657.85 614.34 578.82 522.28 457.76 393.24 329.21 | y* 1297.68 1210.64 1139.61 1026.52 897.48 768.44 640.38 527.29 | # 12 11 10 9 8 7 6 5 4 | aks | | | |
| n 1 2 3 4 5 5 7 3 9 0 | Score Immon. 86.10 60.04 44.05 86.10 102.06 102.06 101.07 86.10 101.07 101.07 | a 86.10 173.13 244.17 357.25 486.29 615.34 743.39 856.48 984.54 1112.60 | atches a ⁺⁺ 43.55 87.07 122.59 179.13 243.65 308.17 372.20 428.74 492.77 556.80 | (Bold F a* | b 114.09 201.12 272.16 385.25 514.29 643.33 771.39 884.47 1012.53 1140.59 | b/87 fr b ⁺⁺ 57.55 101.07 136.58 193.13 257.65 322.17 386.20 442.74 506.77 570.80 | agment : b* | Seq. I S A L E Q L Q Q Q | y 1314.70 1227.67 1156.63 1043.55 914.51 785.46 657.40 544.32 416.26 | y ⁺⁺ 657.85 614.34 578.82 522.28 457.76 393.24 329.21 272.66 | y* 1297.68 1210.64 1139.61 1026.52 897.48 768.44 640.38 527.29 399.24 | # 12 11 10 9 8 7 6 5 4 3 | aks | | | |
| n 1 2 3 4 5 5 5 7 8 9 | Score Immon. 86.10 60.04 44.05 86.10 102.06 102.06 101.07 86.10 101.07 101.07 | a 86.10 173.13 244.17 357.25 486.29 615.34 743.39 856.48 984.54 1112.60 | atches a ⁺⁺ 43.55 87.07 122.59 179.13 243.65 308.17 372.20 428.74 492.77 556.80 | (Bold F a* | b 114.09 201.12 272.16 385.25 514.29 643.33 771.39 884.47 1012.53 1140.59 | b/87 fr b ⁺⁺ 57.55 101.07 136.58 193.13 257.65 322.17 386.20 442.74 506.77 570.80 | agment : b* | Seq. I S A L E Q L Q Q Q | y 1314.70 1227.67 1156.63 1043.55 914.51 785.46 657.40 544.32 416.26 288.20 | y ⁺⁺ 657.85 614.34 578.82 522.28 457.76 393.24 329.21 272.66 208.63 | y* 1297.68 1210.64 1139.61 1026.52 897.48 768.44 640.38 527.29 399.24 271.18 | # 12 11 10 9 8 7 6 5 4 3 2 | aks | | | |

Now we have a run of 11 y ions and a higher score, above the significance threshold. These are not similar sequences with the same set of mass matches.

| | | | | net Explorer | | | | | | |
|--|--|--|---|-------------------------------|--|--|---|---|--|-----|
| le <u>E</u> dit | ⊻iew Fg | vorites | <u>T</u> ools <u>H</u> elp | | | | | | | |
| Back 🔻 | \Rightarrow \cdot \otimes | 1 | 🛛 🧟 Search | 😹 Favorites | History | B- 6 |) 🖬 🖹 | Pow | rmarks 膬 \land 🎋 | |
| iress 🙋 | http://www | .matrixscie | nce.com/cgi/i | master_results. | pl?file=/data/2 | :0000828/ | FTmcobu.da | t | | • 🔊 |
| uery | Ubserv | red B | ir(expt) | Mr(calc |) Delta | Miss | Score | Rank | Peptide | |
| 51 | 1057.0 | 05 2 | 2112.09 | 2112.13 | -0.04 | 0 | 111 | 1 | ALMLQGVD ILAD AVAVTMGPK | |
| <u>52</u> | 1065.0 | 04 2 | 2128.06 | 2128.13 | -0.06 | 0 | (60) | 1 | ALMLQGVDILADAVAVTMGPK + Oxidation (M) | |
| <u>53</u> | | | | 2128.13 | | | (27) | 1 | ALMLQGVDILADAVAVTMGPK + Oxidation (M) | |
| <u>54</u> | 1073.0 | 05 2 | 144.08 | 2144.12 | -0.04 | 0 | (61) | 1 | ALMLQGVDILADAVAVTMGPK + 2 Oxidation (M) | |
| zd9 | | | _fetal_h | | | apiens | | | c hed: 1 IGE:357351 5' similar to gb:X14487_ | |
| Query | | | |) Mr(cal | | | s Score | Rank | Peptide | |
| 33 11- | 714 | . 36 | 1426.71 | 1426.7 | 8 -0.0 | 5 0 | 83 | 1 | ISALEEQLQQIR | |
| 173 ms 234 v1 264 vs 285 vv 405 ms 405 | 82.6 82.6 54.9 54.5 50.2 44.6 44.1 41.7 40.2 | Delta -0.06 -0.10 -0.15 -0.14 0.01 -0.10 -0.17 -0.05 -0.08 -0.20 | | | 1422899 | ISALEEN GTLEKN IASLEQN TGLEEC SISKKIC SLAEIKN SSPEKJ | QLQQIR KLQQIR NQKKLR KLEKRI GALHTWT QEEPRI KLKELR ARTWPR | Ched: IA clo itched IA clo itched IA clo ched: IA clo ched: IA clo ched: IA clo ched: | E IMAGE:607989 5' similar to gb:L0 1 E IMAGE:975759 5' similar to gb:L0 1 E IMAGE:1151658 5' similar to gb:L 1 E IMAGE:1229783 5' similar to gb:L 1 1 E IMAGE:607989 5' similar to gb:L0 1 | |
| | 3b10.r1 329 | Strata 1 Soares | gene mou: Mass: 149 mouse p: | 970 Tota 3NMF19.5 M | 937313) M l score: Mus muscul | us mus B1 Pe j us cDN. | - culus cl p tides r A clone | natched DNA clo natched IMAGE: | ne IMAGE:1150459 5' similar to gb:L 1 173792 5' similar to gb:L00193 Mous | |
| | | | Mass: 17 | ouz Tota | 1 score: | | | | | |
| mi89 | | Soares | mouse en | mbryo NbME | 13.5 14.5 | Mus m | usculus | CDNA C | one IMAGE:479475 5' similar to gb: | • |

The threshold was high because this was an EST search. However, this doesn't change the fact that many people would accept the first match until shown the second.

Our subjective judgement can be misleading

I'm not suggesting that Mascot is infalible, far from it. However, if you choose to disregard the score, you should look very carefully at the match and, ideally, have some additional evidence for it being correct.

6. Peak detection, peak detection, peak detection

- Especially critical for Peptide Mass Fingerprints
- Time domain summing of LC-MS/MS data is very important
- Throw out low mass precursors in MS/MS

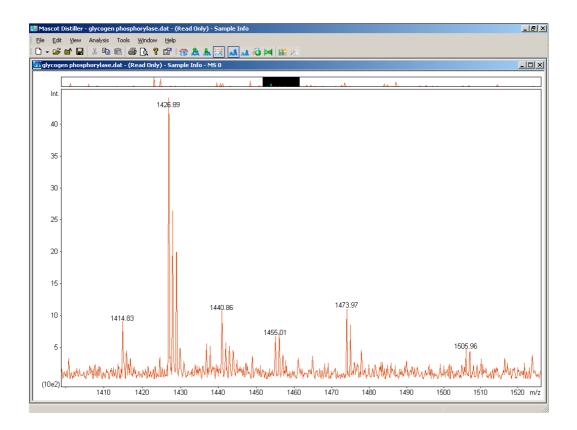
ASMS 2003



If you ask an estate agent (a realtor) in the UK what determines the price of property, they'll probably reply location, location, and location.

Well, in many ways, the quality of a Mascot result depends on peak detection, peak detection, peak detection.

<read from slide>

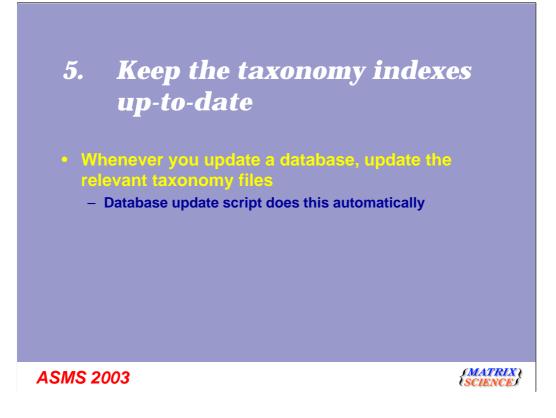


There is a world of difference between a good quality peak list, as you might expect from a piece of software like - random example - Mascot Distiller, and a poor quality peak list, where every spike and glitch on the baseline has been added to the peak list

| | ck 🕶 🗏 |) i 🖉 🖉 🖌 | ථු 📿 Search | 🛐 Favorites 🛛 🕄 | History | à- 🎒 | | Powe | ermarks 🁫 \land 🎋 | |
|------|--|---|--|--|---|--|--|---|--|-----|
| lres | s 🦉 hti | tp://www.matrixs | cience.com/cgi/ma | ster_results.pl?file | =/data/200 | 130602/F | TncCxcn.da | t&REPTYP | E=peptide | • e |
| | | n | | | | | | | | |
| epτ | 1ae 51 | ummary Ro | eport | | | | | | | |
| ritc | h to Pr | otein Summar | y Report | | | | | | | |
| | | | | . ana sa na ana | | | | | | |
| cr | eate a t | bookmark for | this report, rig | ht click this link | : <u>Peptide</u> : | Summar | ry Kepor | t (M. Mo | oss/D. Becherer Sample) | |
| Sel | ect All | Select N | None | Search Selecte | d | Error | tolerant | | | |
| _ | | | | | | | torer and | | | |
| | i 4433 | | Mass: 2 | | 1 score: | 320 | Peptid | les mat | ched: 15 | |
| | | | alin A (Nat | error toles | | ar ou la | | | | |
| Cn | ieck t | o include (| this hit in | error tole | ant sea | ren | | | | |
| Qu | ıery | Observed | Mr(expt) | Mr(calc) | Delta | Miss | Score | Rank | Peptide | |
| | 27 | 959.39 | 958.38 | 958.51 | -0.13 | 0 | (40) | 1 | LLGLFPDAN | |
| | 28 | 480.22 | 958.43 | 958.51 | -0.08 | 0 | 43 | 1 | LLGLFPDAN | |
| | | | | | | | ~~ | | Waawaanaaaaa | |
| | 44 | 659.76 | 1317.50 | 1317.63 | -0.13 | 0 | 80 | 1 | VSSNGSPQGSSVGR | |
| | <u>44</u> 52 | 659.76 524.90 | 1317.50 1571.67 | 1317.63 1571.84 | -0.13 -0.17 | 1 | 80 53 | 1 | VSSNGSPÜGSSVGR VGTAHIIYNSVDKR | |
| | | | | | | | | | | |
| | 52 | 524.90 | 1571.67 | 1571.84 | -0.17 | 1 | 53 | 1 | VGTAHIIYNSVDKR | |
| | <u>52</u> <u>61</u> | 524.90 1051.86 | 1571.67 2101.70 | 1571.84 2102.05 | -0.17 -0.35 | 1 0 | 53 (84) | 1 1 | VGTAHIIYNSVDKR DLILQGDATTGTDGNLELTR | |
| | 52 61 62 | 524.90 1051.86 1051.86 | 1571.67 2101.70 2101.70 | 1571.84 2102.05 2102.05 | -0.17 -0.35 -0.35 | 1 0 0 | 53 (84) (51) | 1 1 1 | VGTAHIIYNSVDKR DLILQGDATTGTDGNLELTR DLILQGDATTGTDGNLELTR | |
| | 52 61 62 63 | 524.90 1051.86 1051.86 1051.86 | 1571.67 2101.70 2101.70 2101.71 | 1571.84 2102.05 2102.05 2102.05 | -0.17 -0.35 -0.35 -0.34 | 1 0 0 0 | 53 (84) (51) (78) | 1 1 1 | VGTAHIIYNSVDKR DLILQGDATTGTDGHLELTR DLILQGDATTGTDGHLELTR DLILQGDATTGTDGHLELTR | |
| | 52 61 62 63 64 | 524.90 1051.86 1051.86 1051.86 1051.86 | 1571.67 2101.70 2101.70 2101.71 2101.71 | 1571.84 2102.05 2102.05 2102.05 2102.05 | -0.17 -0.35 -0.35 -0.34 -0.34 | 1 0 0 0 | 53 (84) (51) (78) (77) | 1 1 1 1 | VGTAHIIYNSVDKR DLILQGDATTGTDGNLELTR DLILQGDATTGTDGNLELTR DLILQGDATTGTDGNLELTR DLILQGDATTGTDGNLELTR | |
| | 52 61 62 63 64 65 | 524.90 1051.86 1051.86 1051.86 1051.86 1051.87 | 1571.67 2101.70 2101.70 2101.71 2101.71 2101.73 | 1571.84 2102.05 2102.05 2102.05 2102.05 2102.05 | -0.17 -0.35 -0.35 -0.34 -0.34 -0.32 | 1 0 0 0 0 | 53 (84) (51) (78) (77) 90 | 1 1 1 1 1 | VGTAHIIYNSVDKR DLILQGDATTGTDGHLELTR DLILQGDATTGTDGHLELTR DLILQGDATTGTDGHLELTR DLILQGDATTGTDGHLELTR DLILQGDATTGTDGHLELTR | |
| | 52 61 63 64 65 66 | 524.90 1051.86 1051.86 1051.86 1051.86 1051.87 701.60 | 1571.67 2101.70 2101.71 2101.71 2101.71 2101.73 2101.78 | 1571.84 2102.05 2102.05 2102.05 2102.05 2102.05 2102.05 | -0.17 -0.35 -0.35 -0.34 -0.34 -0.32 -0.27 | 1 0 0 0 0 0 | 53 (84) (51) (78) (77) 90 (56) | 1 1 1 1 1 1 | VGTAHIITNSVDKR DLILQGDATTGTDGNLELTR DLILQGDATTGTDGNLELTR DLILQGDATGTDGNLELTR DLILQGDATTGTDGNLELTR DLILQGDATGTDGNLELTR DLILQGDATGTDGNLELTR | |
| | 52 61 62 63 64 65 66 67 | 524.90 1051.86 1051.86 1051.86 1051.86 1051.87 701.60 701.60 | 1571.67 2101.70 2101.71 2101.71 2101.71 2101.73 2101.78 2101.79 | 1571.84 2102.05 2102.05 2102.05 2102.05 2102.05 2102.05 2102.05 | -0.17 -0.35 -0.35 -0.34 -0.34 -0.32 -0.27 -0.26 | 1 0 0 0 0 0 0 0 | 53 (84) (51) (78) (77) 90 (56) (52) | 1 1 1 1 1 1 1 | VGTAHIITNSVDKR DLILQGDATTGTDGHLELTR DLILQGDATTGTDGHLELTR DLILQGDATTGTDGHLELTR DLILQGDATTGTDGHLELTR DLILQGDATTGTDGHLELTR DLILQGDATTGTDGHLELTR | |
| | 52 61 62 63 64 65 66 67 68 | 524.90 1051.86 1051.86 1051.86 1051.86 1051.87 701.60 701.60 701.62 | 1571.67 2101.70 2101.71 2101.71 2101.71 2101.73 2101.78 2101.79 2101.82 | 1571.84 2102.05 2102.05 2102.05 2102.05 2102.05 2102.05 2102.05 2102.05 | -0.17 -0.35 -0.35 -0.34 -0.34 -0.32 -0.27 -0.26 -0.23 | 1 0 0 0 0 0 0 0 0 | 53 (84) (51) (78) (77) 90 (56) (52) (46) | 1 1 1 1 1 1 1 1 | VGTAHITYNSVDKR DLILQGDATTGTDGHLELTR DLILQGDATTGTDGHLELTR DLILQGDATTGTDGHLELTR DLILQGDATTGTDGHLELTR DLILQGDATTGTDGHLELTR DLILQGDATTGTDGHLELTR DLILQGDATTGTDGHLELTR | |
| | 52 61 62 63 64 65 66 67 68 68 69 | 524.90 1051.86 1051.86 1051.86 1051.87 701.60 701.60 701.62 1051.93 | 1571.67 2101.70 2101.70 2101.71 2101.71 2101.73 2101.78 2101.79 2101.82 2101.85 | 1571.84 2102.05 2102.05 2102.05 2102.05 2102.05 2102.05 2102.05 2102.05 2102.05 | -0.17 -0.35 -0.35 -0.34 -0.32 -0.27 -0.26 -0.23 -0.20 | 1 0 0 0 0 0 0 0 0 0 | 53 (84) (51) (78) (77) 90 (56) (52) (46) (67) | 1 1 1 1 1 1 1 1 1 | VGTAHITINSVDKR DLILQGDATTGTDGHLELTR DLILQGDATTGTDGHLELTR DLILQGDATTGTDGHLELTR DLILQGDATTGTDGHLELTR DLILQGDATTGTDGHLELTR DLILQGDATTGTDGHLELTR DLILQGDATTGTDGHLELTR DLILQGDATTGTDGHLELTR | |

In the case of MS/MS data, noise peaks aren't such a problem, because Mascot itertively determined which are signal and which are noise. However, time domain processing of LC-MS/MS data is very important.

This example shows what you don't want to see - the same peptide found over and over again. If all these spectra could be summed together, the signal to noise, and hence the Mascot score, would be greatly improved



At number 5, one for the Administrators of in-house Mascot servers: Keep your taxonomy indexes up-to-date <read from slide>

| <pre>iback</pre> | <u>File Edit View Favorites Tools Help</u> | | | |
|--|---|----------------|---------|----------|
| <pre>2#ses@ http://www.matrixsdence.com/v-cg/me-status.exe?0 WLASCOT search status page Fersion 1.9.05a - Jing 5 nodes and 10 processors. [2 searches running] Search log_monitor log_error log_Error message descriptions modelst htt Do not auto refresh this page ame = NCBInr Family = / home/matrix/site/sequence/NCBInr/curre 3 172665 ame = NCBInr Z0030516.fastaPathname = / home/matrix/site/sequence/NCBInr/curre 4 57360 asta 5 27025 6 16074 5 27025 6 16074 5 27025 6 16074 5 27025 6 16074 7 6549 8 3663 9 2013 1 1 274 1 886 1 2744 1 886 1 2744 1 886 1 2744 1 886 1 2744 1 886 1 2744 1 886 1 2744 1 3 511 1 4 470 ame = OWL Family = /home/matrix/site/sequence/NCBInr/curre 1 886 1 2744 1 5 511 1 4 470 ame = OWL Family = /home/matrix/site/sequence/OWI/current/OWL_31.4.fasta 5 tatus 1 In use 5 tatistics Unidentified taxonomy 5 tate Time = Sun Jun 1 04:29:02 # searches = 0 emmapped = YES Request to mem map = YES Request unmap = N0 Mem locked = N0 1 asta 5 tatus = In use 5 tatistics Unidentified taxonomy 1 4 470 ame = OWL Family = /home/matrix/site/sequence/OWI/current/OWL_31.4.fasta 1 1 asta 5 tatus 2 In use 5 tatistics Unidentified taxonomy 5 tate Time = Sun Mun 2 515:47:34 # searches = 0 2 mmapped = YES Request to mem map = YES Request unmap = N0 Mem locked = N0 1 asta 5 tatus 2 In use 5 tatistics Unidentified taxonomy 5 In use 5 tatistics Unidentified taxonomy 5 tate Time = Sun Mun 2 515:47:34 # searches = 0 2 mmapped = YES Request to mem map = YES Request unmap = N0 Mem locked = N0 1 asta 5 tatus 5 In use 5 tatistics Unidentified taxonomy 5 In use 5 In use 5 tatistics Unidentified taxonomy 5 In use 5 In use 5 Tatistics Unidentified taxonomy 5 In use 5 In u</pre> | | | | |
| AASCOT search status page Fersion 1.9.05a- Jang 5 nodes and 10 processors. [2 searches running] Search log monitor log error log Error message descriptions inodelist hat Do not auto refresh this page ame = NCBInr Family = /home/matrix/site/sequence/NCBInr/currer ilename = NCBInr_20030516.fastaPathname = /home/matrix/site/sequence/NCBInr/currer ites Sun Jun 1 04:29:00 # searches = 0 em mapped = NO ER received to mem map = YES Request unmap = NO Mem locked = NO umber of threads = 1 Current = NO ame = NCBInr_20030530.fastaPathname = /home/matrix/site/sequence/NCBInr/currer ites Sun Jun 1 04:29:00 # searches = 0 Estimation = NCBInr_20030530.fastaPathname = /home/matrix/site/sequence/NCBInr/currer ites Sun Jun 1 04:29:00 # searches = 0 Estimation = NCBInr_20030530.fastaPathname = /home/matrix/site/sequence/NCBInr/currer ites Sun Jun 1 04:29:02 # searches = 0 Estimation = NCBInr_20030530.fastaPathname = /home/matrix/site/sequence/NCBInr/currer ites Sun Jun 1 04:29:02 # searches = 0 Estimation = 1 Current = YES mame = OWL = family = /home/matrix/site/sequence/OWI/current/OWL_*.fasta iterame = OWL = family = /home/matrix/site/sequence/OWI/current/OWL_*.fasta iterame = Sun Jun 25 15:47:34 # searches = 0 Estimation = 1 Current = YES mame = NEDB = Family = /home/matrix/site/sequence/MSDB/current/MSDB_*.fasta iterame = MSDB_20030428.fasta Pathname = /home/matrix/site/sequence/MSDB/current/MSDB_*.fasta iterame = MSDB_20030428.fasta Pathname = /home/matrix/site/sequence/MSDB/current/MSDB_*.fasta iterame = MSDB_20030428.fasta Pathname = /home/matrix/site/sequence/MSDB/current/MSDB_20030428.fasta iterame = MSDB_20030428.fasta Pathname = /home/matrix/site/sequence/MSDB/current/MSDB_20030428.fasta iter | | | | |
| <pre>Farson 1.9.05a - Sing 5 nodes and 10 processors. [2 searches running] Search log monitor log error log Error message descriptions nodelist tot Do not auto refresh this page ame = NCBInr</pre> | ddress 🙋 http://www.matrixscience.com/x-cgi/ms-status.exe?0 | | | <u> </u> |
| <pre>Farson 1.9.05a - Sing 5 nodes and 10 processors. [2 searches running] Search log monitor log error log Error message descriptions nodelist tot Do not auto refresh this page ame = NCBInr</pre> | | | | |
| <pre>Tang 5 nodes and 10 processors. [2 searches running] Starch log monitor log error log Error message descriptions nodelist tat Do not auto refresh this page Tax IDs Count Tax IDs C</pre> | MASCOT search status page | | | |
| <pre>Tang 5 nodes and 10 processors. [2 searches running] Starch log monitor log error log Error message descriptions nodelist tat Do not auto refresh this page Tax IDs Count Tax IDs C</pre> | I U | | | |
| Sarch log monitor log error log Error message descriptions modelist tat Do not auto refresh this page Tax IDs Count Sarch log monitor log error log Error message descriptions modelist tat Do not auto refresh this page Tax IDs Count amme = MCBInr Family = /home/matrix/site/sequence/NCBInr/curr 1816669 ilename = NCBInr_20030516.fastaPathname = /home/matrix/site/sequence/NCBInr/curr 172605 tatus = Not in use Statistics Unidentified taxonomy 5 27025 tate Time = Sun Jun 1 04:29:00 # searches = 0 6 16074 meme = NCBInr_20030530.fastaPathname = /home/matrix/site/sequence/NCBInr/curr 9 2013 ilename = NCBInr_20030530.fastaPathname = /home/matrix/site/sequence/NCBInr/curr 10 1274 ilename = NCBInr_20030530.fastaPathname = /home/matrix/site/sequence/NCBInr/curr 10 1274 itename = NCBInr_20030530.fastaPathname = /home/matrix/site/sequence/NCBInr/curr 10 1274 itename = NCBInr_20030530.fastaPathname = /home/matrix/site/sequence/Oul/current/oul_*.fasta 11 8866 itename = Sun Jun 1 04:29:02 # searches = 0 11 866 mame = OWL | /ersion: 1.9.05a - | | | |
| Search tog monator | Jsing 5 nodes and 10 processors. [2 searches running] | | | |
| <pre>dif</pre> | Search log monitor log error log Error message descriptions nodelist txt. Do not auto refresh this page | | nt | |
| <pre>ame = NCBINT</pre> | | | - | |
| <pre>ame = NCBInrFamily = /home/matrix/site/sequence/NCBInr/curred 3 172665 illename = NCBInr_20030516.fastaPathname = /home/matrix/site/sequence/NCBInr/curred 4 57360 fasta tatus = Not in use</pre> | | | | |
| <pre>ilename = NCBInr_20030516.fastaPathame = /home/matrix/site/sequence/NCBInr/curr tatus = Not in use Statistics Unidentified taxonomy tatus = Not in use Statistics Unidentified taxonomy immber of threads = 1 Current = NO mamber of threads = 1 Current = NO mamber of threads = 1 Current = NO = NCBInrFamily = /home/matrix/site/sequence/NCBInr/current ilename = NCBInr_20030530.fastaPathame = /home/matrix/site/sequence/NCBInr/current tatus = In use Statistics Unidentified taxonomy immber of threads = 1 Current = NO = NCBInr_20030530.fastaPathame = /home/matrix/site/sequence/NCBInr/current tatus = In use Statistics Unidentified taxonomy immer of threads = 1 Current = YES = Nome/matrix/site/sequence/Owl/current/Owl_*.fasta ilename = OWLFamily = /home/matrix/site/sequence/Owl/current/owl_*1.fasta ilename = OWLFamily = /home/matrix/site/sequence/Owl/current/Owl_*.fasta ilename = OWLFamily = /home/matrix/site/sequence/Owl/current/Owl_*.fasta ilename = NIBB_20030428.fasta Pathame = /home/matrix/site/sequence/NSDB/current/MSDB_*.fasta ilename = MSDB_20030428.fasta Pathame = /home/matrix/site/sequence/MSDB/current/MSDB_*.fasta ilename = MSDB_20030428.fasta Pathame = /home/matrix/site/sequence/MSDB/current/MSDB_20030428.fasta tatus = In use Statistics Unidentified taxonomy tate Time = Sun Jun 1 22:35:02 # searches = 0 emmapped = YES Request to mem map = YES Request unmap = NO Mem locked = YES = mamped = YES Request to mem map = YES Request unmap = NO Mem locked = YES = mamped = YES Request to mem map = YES Request unmap = NO Mem locked = NO umber of threads = 1 Current = YES = mamped = YES Request to mem map = YES Request unmap = NO Mem locked = YES = mamped = YES Request to mem map = YES Request unmap = NO Mem locked = YES = mamped = YES Request to mem map = YES Request unmap = NO Mem locked = YES = mamped = YES Request to mem map = YES Request unmap = NO Mem locked = YES = mamped = YES Request to mem map = YES Request unmap = NO Mem locked = YES = Mamper SES Request to mem map = YES Request unm</pre> | ame = NCBInr Family = /home/matrix/site/sequence/NCBInr/curre | | | |
| <pre>tatus = Not in use Statistics Unidentified taxonomy 5 27025 tate Time = Sun Jun 1 04:29:00 # searches = 0 tate Time = Sun Jun 1 04:29:00 # searches = 0 tate Time = NO Request to mem map = YES Request unmap = NO Mem locked = NO tatus = In use</pre> | | 5 114 | | |
| <pre>tate Time = Sun Jun 1 04:29:00 # searches = 0 immapped = NO Request to mem map = YES Request unmap = NO Mem locked = NO 7 6349 ame = NCBInr</pre> | | 4 575 | | |
| <pre>lem mapped = NO Request to mem map = YES Request unmap = NO Mem locked = NO</pre> | tate Time = Sun Jun 1 04:29:00 # searches = 0 | | | |
| <pre>umber of threads = 1 Current = NO ame = NCBInrFamily = /home/matrix/site/sequence/NCBInr/curre ilename = NCBInr_20030530.fastaPathnee = /home/matrix/site/sequence/NCBInr/curre il 88</pre> | | | | |
| <pre>ame = NCEInr Family = /home/matrix/site/sequence/NCEInr/currer 9 2013 11ename = NCEInr 20030530.fastaPathapene = /home/matrix/site/sequence/NCEInr/currer 10 1274 1asta 1274 1asta 1274 1asta 1274 1asta 12 744 13 511 14 470 ame = OWL Family = /home/matrix/site/sequence/Owl/current/owl *.fasta 11 ename = OWL</pre> | umber of threads = 1 Current = NO | | | |
| <pre>ame = NCBIN_20030530.fastBarthnew = /home/matrix/site/sequence/NCBIN/current/10 1274 tatus = In use Statistics Pridentified taxonomy tate Time = Sun Jun 1 04:29:02 # searches 0</pre> | | | | |
| <pre>intername = NSDE_200305201285102 # searches = 0 ame = MSDE_20030428.fasta Pathname = /home/matrix/site/sequence/MSDE/current/MSDE_20030428.fasta tatus = In use Statistics Unidentified taxonomy tate Time = Sun Jun 1 04:29:02 # searches = 0 ame = OWL</pre> | | | | |
| <pre>Solution = 10 Note = 10 Section = 10 Se</pre> | | | - Iasta | |
| <pre>iem mapped = YES Request to mem map = YES Request unmap = NO Mem locked = YES imber of threads = 1 Current = YES imber of threads = 1 Current = YES ili 13 511 14 470 iliename = Owl_31.4.fasta Pathname = /home/matrix/site/sequence/Owl/current/Owl_*.fasta iliename = Owl_31.4.fasta Pathname = /home/matrix/site/sequence/Owl/current/Owl_31.4.fasta tatus = In use Statistice Unidentified taxonomy tate Time = Sun May 25 15:47:34 # searches = 0 iem mapped = YES Request to mem map = YES Request unmap = NO Mem locked = NO imber of threads = 1 Current = YES imber of threads = YES imber of threads = YES imber of</pre> | | 12 744 | | |
| <pre>umber of threads = 1 Current = YES 14 470 ame = OWL Family = /home/matrix/site/sequence/Owl/current/Owl*.fasta ilename = Owl_31.4.fasta Pathname Statistics Unidentified taxonomy tate Time = Sun May 25 15:47:34 # searches = 0 em mapped = YES Request to mem map = YES Request unmap = N0 Mem locked = N0 umber of threads = 1 Current = YES ame = MSDB_20030428.fasta Pathname = /home/matrix/site/sequence/MSDB/current/MSDB_*.fasta tatus = In use Statistics Unidentified taxonomy ame = MSDB_20030428.fasta Pathname = /home/matrix/site/sequence/MSDB/current/MSDB_*.fasta tatus = In use Statistics Unidentified taxonomy tate Time = Sun Jun 1 22:35:02 # searches = 0 em mapped = YES Request to mem map = YES Request unmap = N0 Mem locked = YES</pre> | | 13 511 | | |
| <pre>ilename = 0vl_31.4.fasta Pathname = /home/matrix/site/sequence/0vl/current/0vl_31.4.fasta tatus = In use Statistics Unidentified taxonomy tate Time = Sun May 25 15:47:34 # searches = 0 em mapped = YES Request to mem map = YES Request unmap = NO Mem locked = NO umber of threads = 1 Current = YES ame = MSDB_20030428.fasta Pathname = /home/matrix/site/sequence/MSDB/current/MSDB_*.fasta ilename = MSDB_20030428.fasta Pathname = /home/matrix/site/sequence/MSDB/current/MSDB_20030428.fasta tatus = In use Statistics Unidentified taxonomy tate Time = Sun Jun 1 22:35:02 # searches = 0 em mapped = YES Request to mem map = YES Request unmap = NO Mem locked = YES</pre> | Number of threads = 1 Current = YES | 14 470 | | |
| <pre>ilename = 0vl_31.4.fasta Pathname = /home/matrix/site/sequence/0vl/current/0vl_31.4.fasta tatus = In use Statistics Unidentified taxonomy tate Time = Sun May 25 15:47:34 # searches = 0 em mapped = YES Request to mem map = YES Request unmap = NO Mem locked = NO umber of threads = 1 Current = YES ame = MSDB_20030428.fasta Pathname = /home/matrix/site/sequence/MSDB/current/MSDB_*.fasta ilename = MSDB_20030428.fasta Pathname = /home/matrix/site/sequence/MSDB/current/MSDB_20030428.fasta tatus = In use Statistics Unidentified taxonomy tate Time = Sun Jun 1 22:35:02 # searches = 0 em mapped = YES Request to mem map = YES Request unmap = NO Mem locked = YES</pre> | | | | |
| <pre>ilename = 0vl_31.4.fasta Pathname = /home/matrix/site/sequence/0vl/current/0vl_31.4.fasta tatus = In use Statistics Unidentified taxonomy tate Time = Sun May 25 15:47:34 # searches = 0 em mapped = YES Request to mem map = YES Request unmap = NO Mem locked = NO umber of threads = 1 Current = YES ame = MSDB_20030428.fasta Pathname = /home/matrix/site/sequence/MSDB/current/MSDB_*.fasta ilename = MSDB_20030428.fasta Pathname = /home/matrix/site/sequence/MSDB/current/MSDB_20030428.fasta tatus = In use Statistics Unidentified taxonomy tate Time = Sun Jun 1 22:35:02 # searches = 0 em mapped = YES Request to mem map = YES Request unmap = NO Mem locked = YES</pre> | | | | |
| <pre>tates = In use Statistics Unidentified taxonomy tate Time = Sun May 25 15:47:34 # searches = 0 enmapped = YES Request to mem map = YES Request unmap = NO Mem locked = NO umber of threads = 1 Current = YES ame = MSDB Family = /home/matrix/site/sequence/MSDB/current/MSDB_*.fasta ilename = MSDB Tame = /home/matrix/site/sequence/MSDB/current/MSDB_20030428.fasta tatus = In use Statistics Unidentified taxonomy tate Time = Sun Jun 1 22:35:02 # searches = 0 enmapped = YES Request to mem map = YES Request unmap = NO Mem locked = YES</pre> | | | | |
| <pre>tate Time = Sun May 25 15:47:34 # searches = 0 'em mapped = YES Request to mem map = YES Request unmap = NO Mem locked = NO umber of threads = 1 Current = YES make = MSDE_20030428.fasta Pathname = /home/matrix/site/sequence/MSDE/current/MSDE_20030428.fasta tatus = In use Statistics Unidentified taxonomy tate Time = Sun Jun 1 22:35:02 # searches = 0 em mapped = YES Request to mem map = YES Request unmap = NO Mem locked = YES</pre> | | 001_31.4.1ast | a | |
| <pre>iem mapped = YES Request to mem map = YES Request unmap = N0 Mem locked = N0 umber of threads = 1 Current = YES umber of threads = 1 Current = YES imme = MSDB</pre> | | | | |
| umber of threads = 1 Current = YES ame = <u>HSDB</u> Family = /home/matrix/site/sequence/HSDB/current/HSDB_*.fasta ilename = MSDB_20030428.fasta Pathname = /home/matrix/site/sequence/MSDB/current/MSDB_20030428.fasta tatus = In use <u>Statistics Unidentified taxonomy</u> tate Time = Sun Jun 1 22:35:02 # searches = 0 en mapped = YES Request to mem map = YES Request unmap = NO Mem locked = YES | | | | |
| ilename = MSDB_20030428.fasta Pathname = /home/matrix/site/sequence/MSDB/current/MSDB_20030428.fasta tatus = In use <u>Statistics</u> <u>Unidentified taxonomy</u> tate Time = Sun Jun 1 22:35:02 # searches = 0 em mapped = YES Request to mem map = YES Request unmap = NO Mem locked = YES | Number of threads = 1 Current = YES | | | |
| ilename = MSDB_20030428.fasta Pathname = /home/matrix/site/sequence/MSDB/current/MSDB_20030428.fasta tatus = In use <u>Statistics</u> <u>Unidentified taxonomy</u> tate Time = Sun Jun 1 22:35:02 # searches = 0 em mapped = YES Request to mem map = YES Request unmap = NO Mem locked = YES | | | | |
| ilename = MSDB_20030428.fasta Pathname = /home/matrix/site/sequence/MSDB/current/MSDB_20030428.fasta tatus = In use <u>Statistics</u> <u>Unidentified taxonomy</u> tate Time = Sun Jun 1 22:35:02 # searches = 0 em mapped = YES Request to mem map = YES Request unmap = NO Mem locked = YES | | (WODD + c | | |
| tatus = In use <u>Statistics</u> <u>Unidentified taxonomy</u> | | | | |
| tate Time = Sun Jun 1 22:35:02 | | 7 nau6_2003042 | o.Iasta | |
| em mapped = YES Request to mem map = YES Request unmap = NO Mem locked = YES | | | | |
| | | | | |
| | | | | |
| | | | | |

From time to time, its a good idea to check the stats file for each database. It contains lots of useful information, like whether entries contain illegal characters or whether an entry is too long.

It also tells you how good your taxonomy is. Here are the numbers for the nr database on our web site at the end of May. There are 1.4 million entries, but only 1200 have no taxonomy. In other words, better than 99.9% of the entries have a taxonomy assigned. If you look at your stats file and see that (say) 10% of the entries have no taxonomy, that's 10% of the entries that are going to be missed whenever you do a search with taxonomy specified.

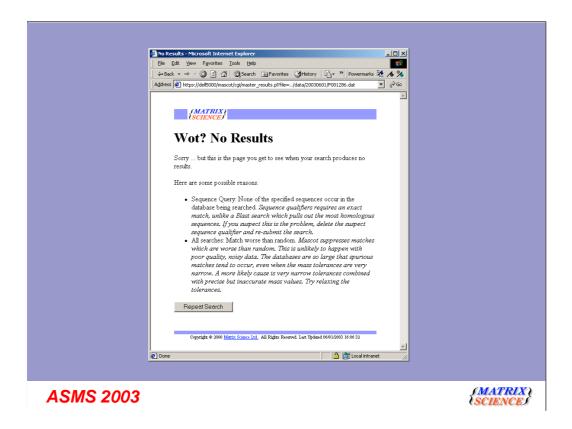
4. Remember that enzyme specificity also applies to Sequence Queries

ASMS 2003

Top tip number 4 is Remember that enzyme specificity also applies to Sequence Queries

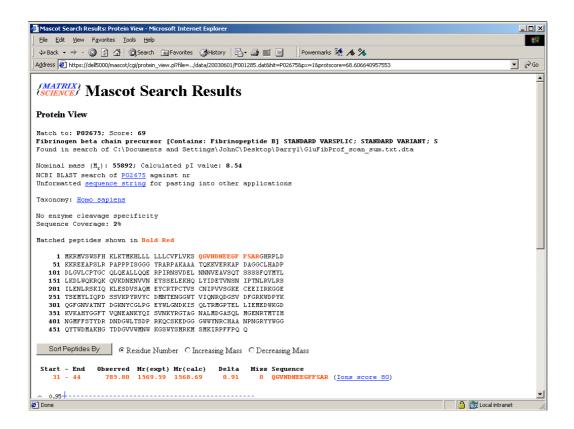
SCIENCES

One of the most common emails we receive is "Mascot is broken. I did a search for this peptide and I know its in the database but Mascot failed to find it"



For example, here's a search for glu-fib, a very common sequencing standard. The mass is correct and the sequence is correct. But, when we do a search of Swiss-Prot - No results!

Why



Because glu-fib in Swiss-Prot is not a tryptic peptide. The N-terminus is created by a post-translational cleavage after serine. If you now go back to the search form and select enzyme type none, bingo ... you'll get a match

3. **Don't specify a protein mass** unless essential

- · Cannot guarantee that the mass of the database entry is close to that of the analyte
- Never useful for MS/MS search. Only useful for Peptide Mass Fingerprint when
 - Analyte is small fragment of very large entry
 - Low complexity entry

ASMS 2003



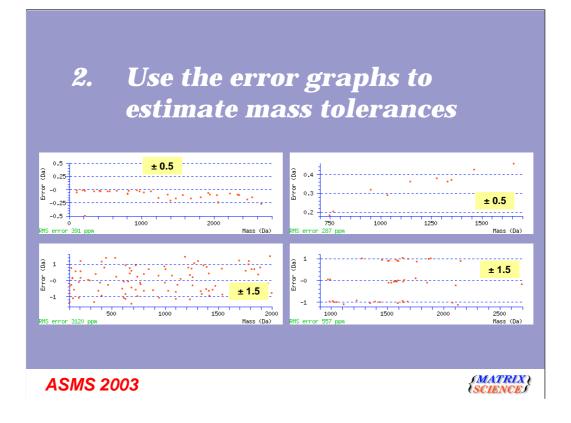
Number 3 is another very common technical support issue: Whether to specify a protein mass

<read from slide>

| e <u>E</u> dil | t ⊻iew In | sert F | ormat | <u>T</u> ools Ta | ble \ | <u>W</u> indow <u>H</u> elp | | | | | | | |
|----------------|-----------|--------|-------|------------------|-------|-----------------------------|----------|--------------------------|------|-------|---------|------|----------|
| | | 117% | | | 1 🔲 |] Close N? | _ | | | | | | |
| | | | _ | | | | <u> </u> | | | | | | |
| - | | | | 1 | | | | | | | | | |
| | | | | | | | | pgvrweycnl | | | sleaps | edap | |
| | | | | | | | | tysttvtgrt | | | | | |
| | | | | | | | | ycnltqcsda | | A | POA | - HI | UMAN: |
| | | | | | | | | tgrtcqawss csdaeqtava | | | | | |
| | | | | | | | | awssmtphsh | | 45 | 48 A/ | A. c | of which |
| | | | | | | | | tavapptvtp | | | | | |
| | | | | | | | | phshsrtpey | | | 4218 | AA | \ is 37 |
| | | | | | | | | tvtpvpslea | | | e e f c | - 5 | - 444 |
| | | | | | | | | tpeyypnagl | | rep | eats | OT | a 114 AA |
| | | | | | | | | sleapsegap | | | | | |
| | | | | | | | | naglimnycr | | r | \ringi | le o | lomain |
| | | | | | | | | eqapteqrpq | | t var | | | |
| | | | | | | | | nycrnpdava | | | | | |
| | | | | | | | | drpgvqecyh | | | | | |
| | | | | | | | | davaapycyt | | | | | |
| | | | | | | | | ecyhgngqsy | | | | | |
| | | | | | | | | ycytrdpgvr | | | | | |
| | | | | | | | | qqsyrqtyst | | | | | |
| | | | | | | | | pqvrweycnl | | | | | |
| | | | | | | | | tysttvtqrt | | | | | |
| | | | | | | | | vcnltqcsda | | | | | |
| | | | | | | | | tgrtcgawss | | | | | |
| | | | | | | | | csdaeqtava | | | | | |
| | | | | | | | | awssmtphsh | | | | | |
| | | | | | | | | tavapptvtp | | | | | |
| | | | | | | | | phshsrtpey | | | | | |
| | | | | | | | | tvtpvpslea | | | | | |
| | | | | | | | | tpeyypnagl | | | | | |
| | | | | | | | | sleapseqap | | | | | |
| | | | | | | | | naglimnycr | | | | | |
| | | | | | | | | egaptegrpg | | | | | |
| | | | | | | | | nycrnpdava | | | | | |
| | | | | | | | | grpgvgecyh | | | | | |
| | | | | | | | | davaapycyt | | | | | |
| | | | | | | | | ecyhgngqsy | | | | | |
| G 🗉 | ⊡ 4 | | | | | | | | | - | - | | |

Here, for example, is human Apolipoprotein a Almost all of this protein is a repeated kringle domain of just 114 residues. Statistically, this protein behaves like a much smaller protein ... for eample, it will produce many fewer unique tryptic peptides than you would expect from its size. If you had a peptide mass map of this protein, it would be very, very difficult to get a match without specifiying a small protein mass.

This, and the case where the experimental protein is a very small fragment of the database entry are the times you need to use SEG. Otherwise, much better to leave the protein mass open



Number 2 is a reminder to use the error graphs to estimate mass tolerances

1. This example is fine, the mass errors are well within the specified tolerance of +/- 0.5. You could probably increase the score slightly by going to +/- 0.3, but safer to leave it where it is

2. This is also fine! The mass values are mostly within the specified tolerance of +/-1.5. In fact, this is the error distribution for a very good MS/MS match from an ion trap.

3. In contrast, this is not right. Although the accuracy is better than the last example, the mass scale should continue to 2500 Da. However, all the potentail matches above 1650 Da have been lost because the tolerance is too tight and is clipping the high masses. The precision suggests that some calibration is overdue

4. This is a worrying example. The accuracy is excellent, but a very wide tolerance has been specified. For a peptide mass fingerprint, this can easily create a false positive, because the distribution of mass values is is not uniform. This kind of data is playing with Mascot's mind. I don't have time to go into great detail. Suffice to say that if you see this, you should set a more appropriate tolerance, like +/- 0.5.

1. Be sparing with variable modifications

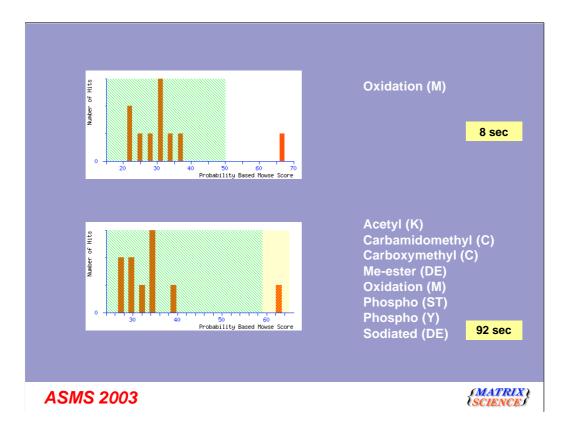
· Some modifications are worse than others

- Mods that affect a terminus are less of a problem, e.g. Pyro-glu
- Mods that apply to residue(s) with a high fractional abundance and at any position are BIG problem, e.g. Phospho (ST) = 13%
- Use an error tolerant search to pick up
 uncommon modifications
 - Efficient
 - Also catch non-specific peptides

ASMS 2003



And finally, number 1, our top tip! Be sparing with variable modifications <read from slide>



This search of a single MS/MS spectrum, using one variable mod, gives a nice, statistically significant match.

If the search is repeated with $8 \ \text{mods},$ the match is the same, but it is no longer so clear cut.

All of these mods have effectively increased the size of the database by a factor of $30!\,$

Whats worse, the search takes over 10 times as long!

So, our top tip is to use variable mods sparingly. You'll get better results faster.

- 1. Be sparing with variable modifications
- 2. Use the error graphs to estimate mass tolerances
- 3. Don't specify a protein mass unless essential
- 4. Remember that enzyme specificity also applies to Sequence Queries
- 5. Keep the taxonomy indexes up-to-date
- 6. Peak detection, peak detection, peak detection
- 7. Be skeptical if Mascot score is below threshold
- 8. Submit new modifications to Unimod
- 9. Use the Peptide Summary Report for MS/MS results
- 10. Don't specify a poorly represented taxonomy

ASMS 2003

