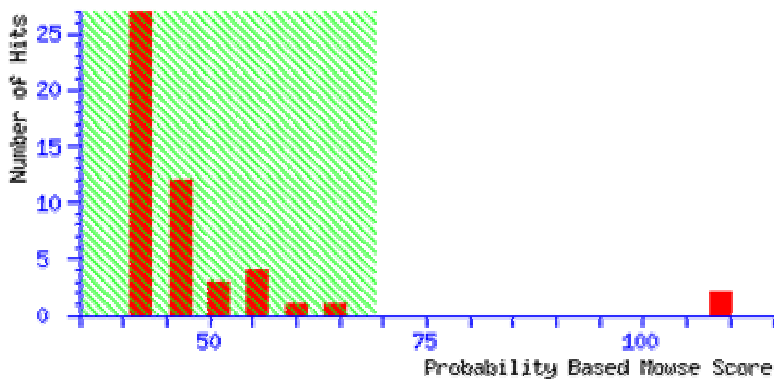

Mascot Lab: ABRF Workshop

Feb. 5, 2005, Savannah, Georgia, USA

David Wishart, University of Alberta, Edmonton, Alberta, Canada

Introduction

Mascot is a web-based tool for predicting protein sequences from PMF and MS/MS data. The purpose of this lab is to familiarize you with the basic use of Mascot and the interpretation of Mascot results.



Example 1: Mascot Peptide Mass Fingerprint Analysis

A yeast fractionated sample has been treated with iodoacetamide (carbamidomethylation of the Cysteine residues) and then with 1-D SDS electrophoresis. One of the bands (about 75KDa) has been submitted to trypsin digestion followed by MALDI-TOF MS measurement.

1. Open Netscape or Internet Explorer and go to <http://gchelpdesk.ualberta.ca/ABRF2005/> and look for the 'example1.txt' link. Click on the link to view the file. Save the file to your home folder.

Mascot Search

- Peptide Mass Fingerprint:** The experimental data are a list of peptide mass values from an enzymatic digest of a protein.
 - ◊ Example of results report
 - ◊ More information
- Sequence Query:** One or more peptide mass values associated with information such as partial or ambiguous sequence strings, amino acid composition information, MS/MS fragment ion masses, etc. A super-set of a sequence tag query.
 - ◊ Example of results report
 - ◊ More information
- MS/MS Ion Search:** Identification based on raw MS/MS data from one or more peptides.
 - ◊ Example of results report
 - ◊ More information

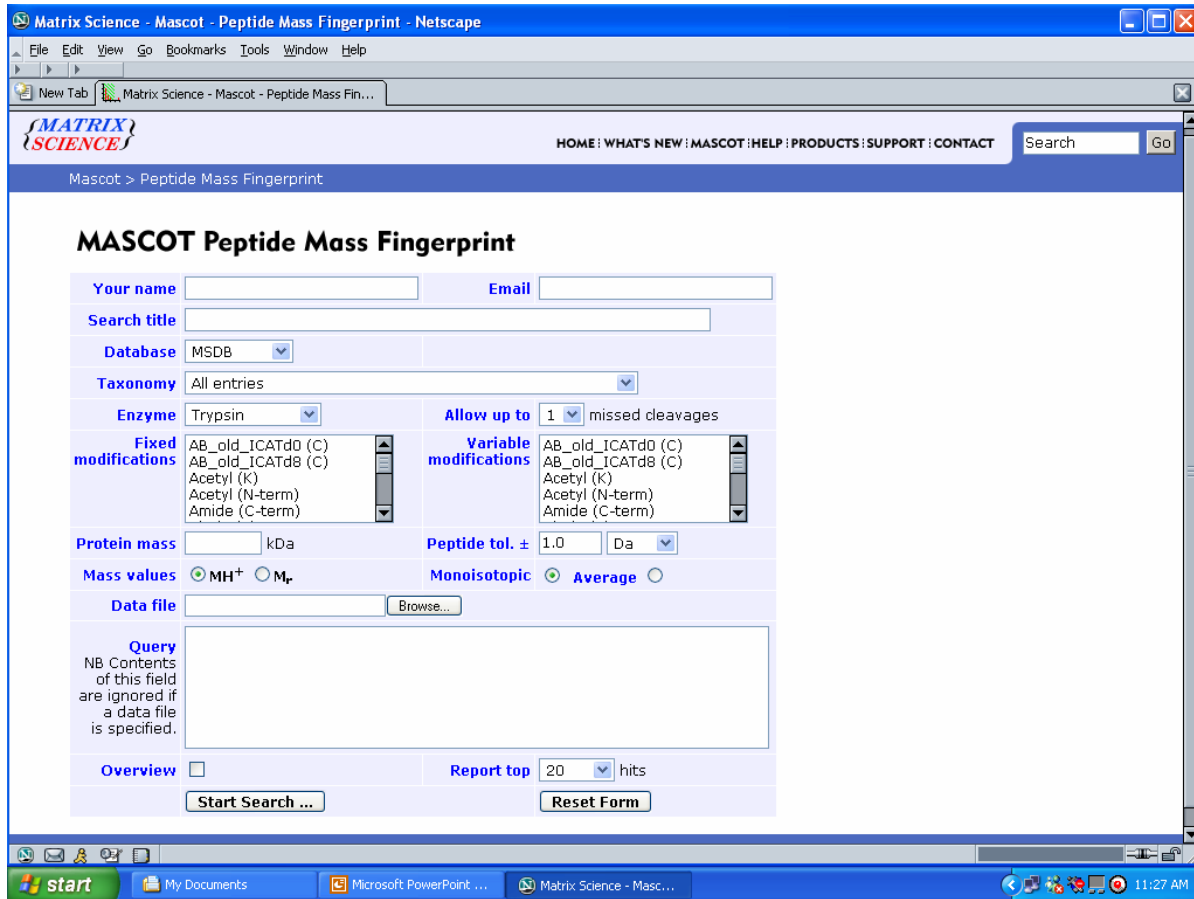
Search Form Defaults: Follow this link to save your preferred search form defaults as a browser cookie.

Mascot Help

- Mascot Overview
- Search parameter reference
- Sequence databases
- Data file format
- Scoring algorithm
- Results format
- Error tolerant search
- FAQ's
- User Meeting Presentations
- 2004
- More Help
- Help Topic Index
- Useful Links

Copyright © 2003 Matrix Science Ltd. All Rights Reserved. Last Updated Thu, 19 Aug 2004 13:04:26 GMT

2. Go to the **Mascot** website (www.matrixscience.com). This takes you to the Matrix Science homepage, which hosts Mascot. Click the '**Mascot**' link on the top of the form to go to the Mascot home page. Select the **Peptide Mass Fingerprint Search Form**. This search uses MALDI-TOF spectra of peptide ion masses from a protein digest to predict the protein sequence.



Fill in the fields as follows:

Your name: fill in with your name. Nicknames are okay too.

Email: your real email address. Mascot will email your results to you if you get disconnected or time out.

Search Title: use 'example 1' or some other descriptive title.

Database: this selects the database for Mascot to search against. Choose the **NCBI**.

Taxonomy: allows the searches to be restricted to a specific species or group of species. Use '**All Species**'.

Enzyme: the cleavage enzyme used to digest the protein into peptide fragments. Choose '**Trypsin**'.

Missed Cleavages: often a sample is only partially digested, resulting in peptides with intact cleavage sites. Increasing the number of missed cleavages increases the probability of

identifying missed cleavages in your sample, but it is computationally very expensive as the number goes beyond 2 or 3. Choose '1' allowed missed cleavage site.

Fixed modifications: this table allows you to select modifications to the amino acids, peptide N- and C-termini, etc. The modification is applied to every instance of the modified group. This sample has been treated with iodoacetamide, so in the fixed modifications box, select '**Carbamidomethyl (C)**'.

Variable Modifications: like the fixed modifications, but Mascot here only 'considers' the modification; unmodified masses are equally considered. Leave the Variable Modifications list unselected.

Protein Mass: use **75** kDa.

Peptide Tolerance: this is a measure of the the mass spectrometer's ability to precisely report the sample peptide masses. Use a tolerance of **+/- 1 Da**.

Mass Values: this option lets you choose whether the masses have been reported with the charged proton included (MH+) or without (Mr). Choose **MH+**.

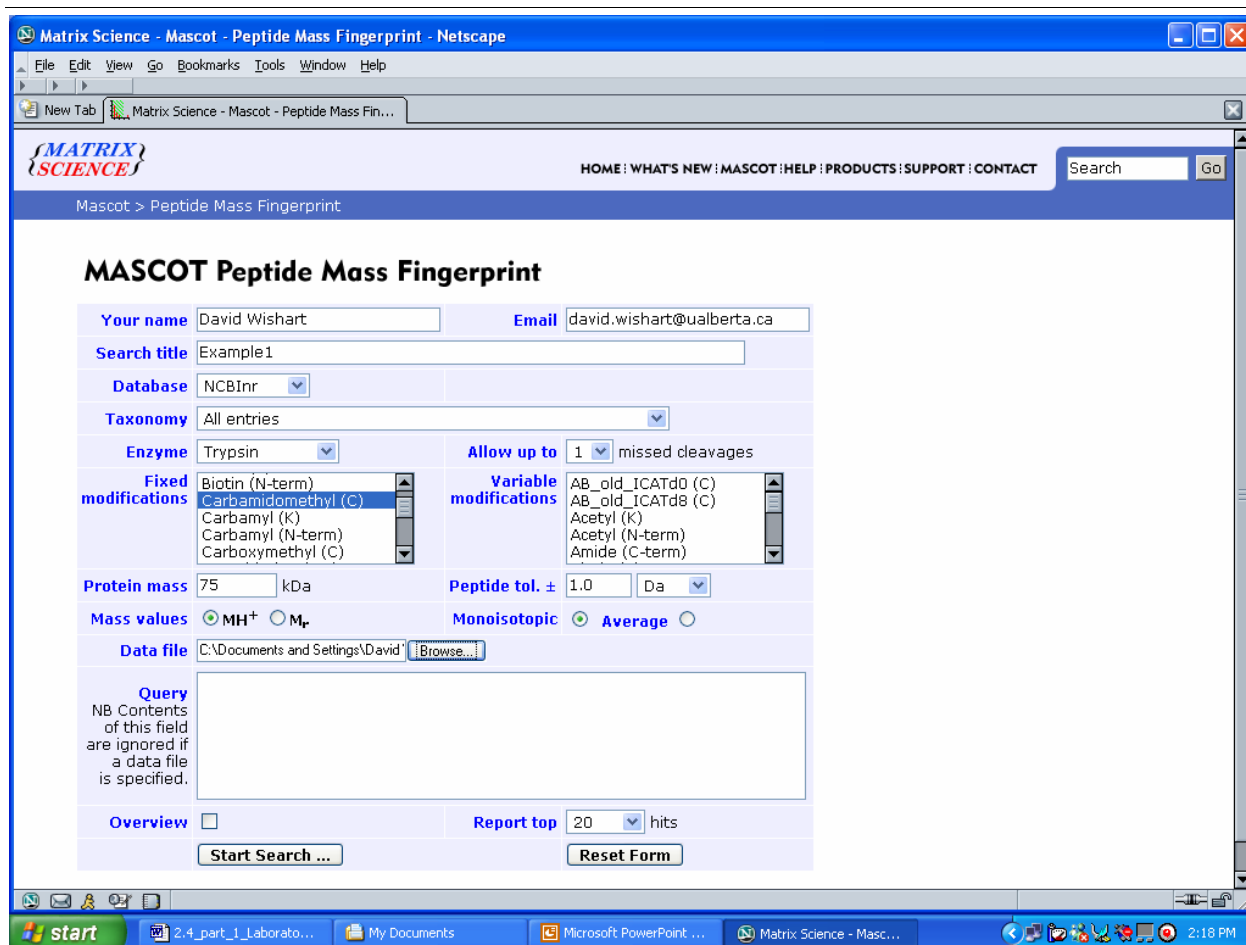
Monoisotopic or Average: The monoisotopic mass is the mass determined using the masses of the most abundant isotopes. The average mass is the abundance weighted mass of all isotopic components. Modern mass spectrometers can readily resolve the various isotope peaks of peptides and proteins, so choose **Monoisotopic**.

Data File: this is the ion peak table. Click Browse, then select the **example1.txt**.

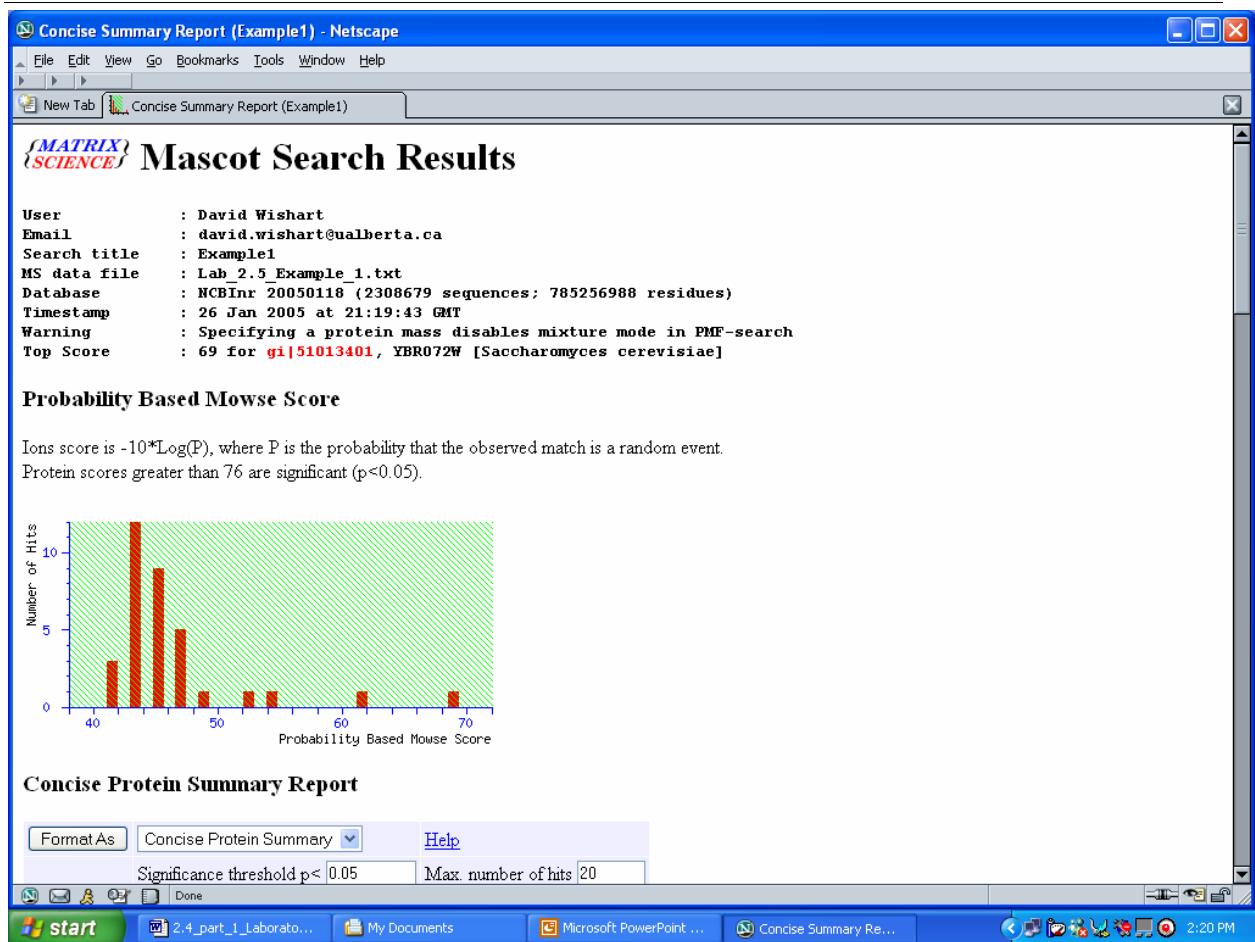
Overview: this adds some eyecandy but it not useful for mass fingerprint analyses, so leave it **unchecked**.

Report top: the number of hits to return. Select **20**.

Your search form should now look like the one provided below.



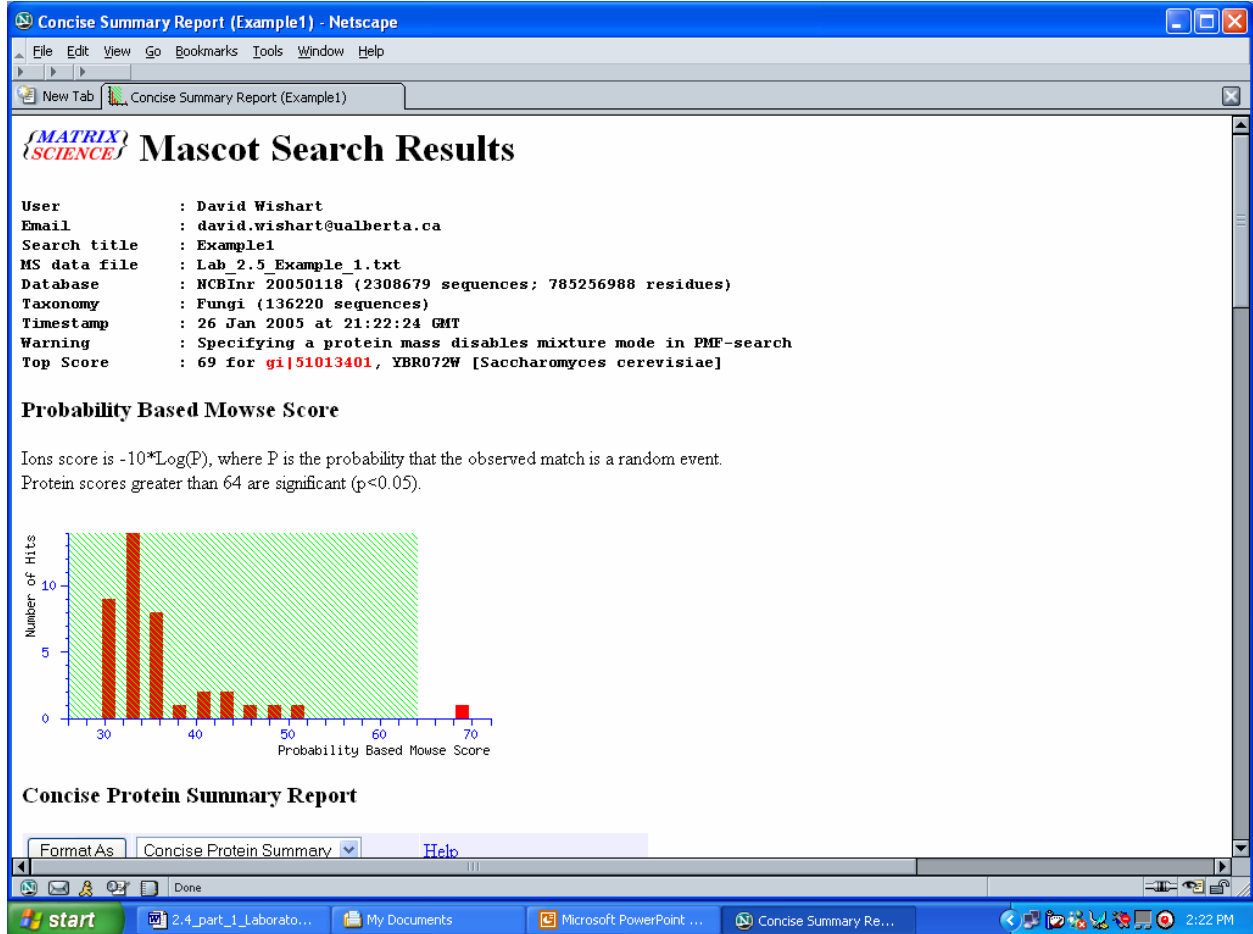
4. **START** the Mascot Search (not all at once: you'll crash the Mascot server!)



Mascot reports a top score of 69 for heat shock protein 26. But is it significant? Not according to Mascot. Mascot reports a match as significant if it has a match with a less than 5% chance of being a random hit. The top scoring matches here do not meet this criteria.

The chance of getting a random hit is proportional to the number of sequences being searched. One way to reduce that chance, and therefore increase the significance of a hit is to reduce the number of sequences being searched by restricting the search to a specific group of species. We know the hit is derived from yeast, so we don't need to search the entire NR database.

5. Go back to the search form and select '**Fungi**' from the **taxonomy** field.
6. Rerun the search.



The Mascot score remains the same as before, but the hit is now significant.

- At the bottom of the window, use the pull down box to change “Concise Protein Summary” to “Protein Summary” and press the “**Format As**” button. You should see a new screen generated on your browser (see below). Scroll down and look at the peptides identified from the top scoring protein.

Protein Summary Report (Example1) - Netscape

Accession	Mass	Score	Description
1. gi 51013401	23736	69	YBR072W [Saccharomyces cerevisiae]
2. gi 295614	23952	69	heat shock protein 26
3. gi 6319546	23865	69	Hsp26p [Saccharomyces cerevisiae]
4. gi 19112530	15499	52	40s ribosomal protein s16. [Schizosaccharomyces pombe]
5. gi 46097766	96068	48	hypothetical protein UM05125.1 [Ustilago maydis 521]
6. gi 44985089	148668	45	AFR496Wp [Ashbya gossypii ATCC 10895]
7. gi 38107647	18468	44	hypothetical protein MG09542.4 [Magnaporthe grisea 70-15]
8. gi 46431898	104432	43	conserved hypothetical protein [Candida albicans SC5314]
9. gi 11276570	15240	41	ribosomal protein S16 - fission yeast (Schizosaccharomyces pombe) (fragment)
10. gi 6470356	167711	40	polyprotein [Candida albicans]
11. gi 6323101	78619	40	Protein of unknown function; green fluorescent protein (GFP)-fusion protein localizes to the
12. gi 46431192	167654	39	retrotransposon Tca5 polyprotein [Candida albicans SC5314]
13. gi 2565014	125059	38	Cdp1p [Saccharomyces cerevisiae]
14. gi 12004595	42746	37	uridine nucleosidase [Saccharomyces cerevisiae]
15. gi 19115207	72740	36	hypothetical protein [Schizosaccharomyces pombe]
16. gi 49653622	134077	36	unnamed protein product [Debaryomyces hansenii CBS767]
17. gi 49647784	54863	35	unnamed protein product [Yarrowia lipolytica CLIB99]
18. gi 38109783	99710	35	hypothetical protein MG01251.4 [Magnaporthe grisea 70-15]
19. gi 19074347	33370	35	hypothetical protein [Encephalitozoon cuniculi]
20. gi 46444202	60971	34	transposase-like protein [Candida albicans SC5314]

Results List

1. [gi|51013401](#) **Mass:** 23736 **Score:** 69 **Expect:** 0.018 **Queries matched:** 7

YBR072W [Saccharomyces cerevisiae]

Observed	Mr(expt)	Mr(calc)	Delta	Start	End	Miss	Peptide
998.60	997.59	997.57	0.03	37 - 45	1		RQLANTPAK
1274.60	1273.59	1273.59	-0.00	118 - 127	0		DIDIEYHQNK
1461.80	1460.79	1460.79	0.01	177 - 190	0		ADYANGVLTITVPE
1805.90	1804.89	1804.89	0.01	51 - 66	0		EVARPNNYAGALYDPR
1886.00	1884.99	1884.99	-0.00	160 - 176	1		RVITLDPYPGVDADNIK
2041.00	2039.99	2040.02	-0.02	90 - 107	0		SVAVPVDILDHDMNYELK
2201.10	2200.09	2200.12	-0.03	128 - 147	1		NQILVSGEIPSTLNESKDK

7. Click on the accession number of the top protein on the "Results List" to see the peptide coverage over the protein. This will generate an image with the complete sequence of the protein and the identified peptides marked off in red as seen on the next page.

Mascot Search Results

Protein View

Match to: [gi|51013401](#) Score: 69 Expect: 0.018
YBR072W [Saccharomyces cerevisiae]
 Found in search of Lab_2.5_Example_1.txt

Nominal mass (M_r): 23736; Calculated pI value: 5.31
 NCBI BLAST search of [gi|51013401](#) against nr
 Unformatted [sequence string](#) for pasting into other applications

Taxonomy: [Saccharomyces cerevisiae](#)

Fixed modifications: Carbamidomethyl (C)
 Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
 Number of mass values searched: 18
 Number of mass values matched: 7
 Sequence Coverage: 48%

Matched peptides shown in **Bold Red**

1 MSFNSPFDFD FDNINNEVDA FNRLLEGGL RGYAPRRQLA NTPAKDSTGK
 51 **EVARPNNYAG ALYDPRDETL DDWFDNDLSL FPSGFGFPRS VAVPVDILDH**
 101 **DNNYELKVVV PGVSKKRDID IEYHQNKQI LVSGEIPSTL NEESKDKVKV**
 151 **KESSSGKFKR VITLDPYGV DADNIKADYA NGVLTITVPK LKPQDKGNH**
 201 VKKIEVSSQE SGMN

Show predicted peptides also

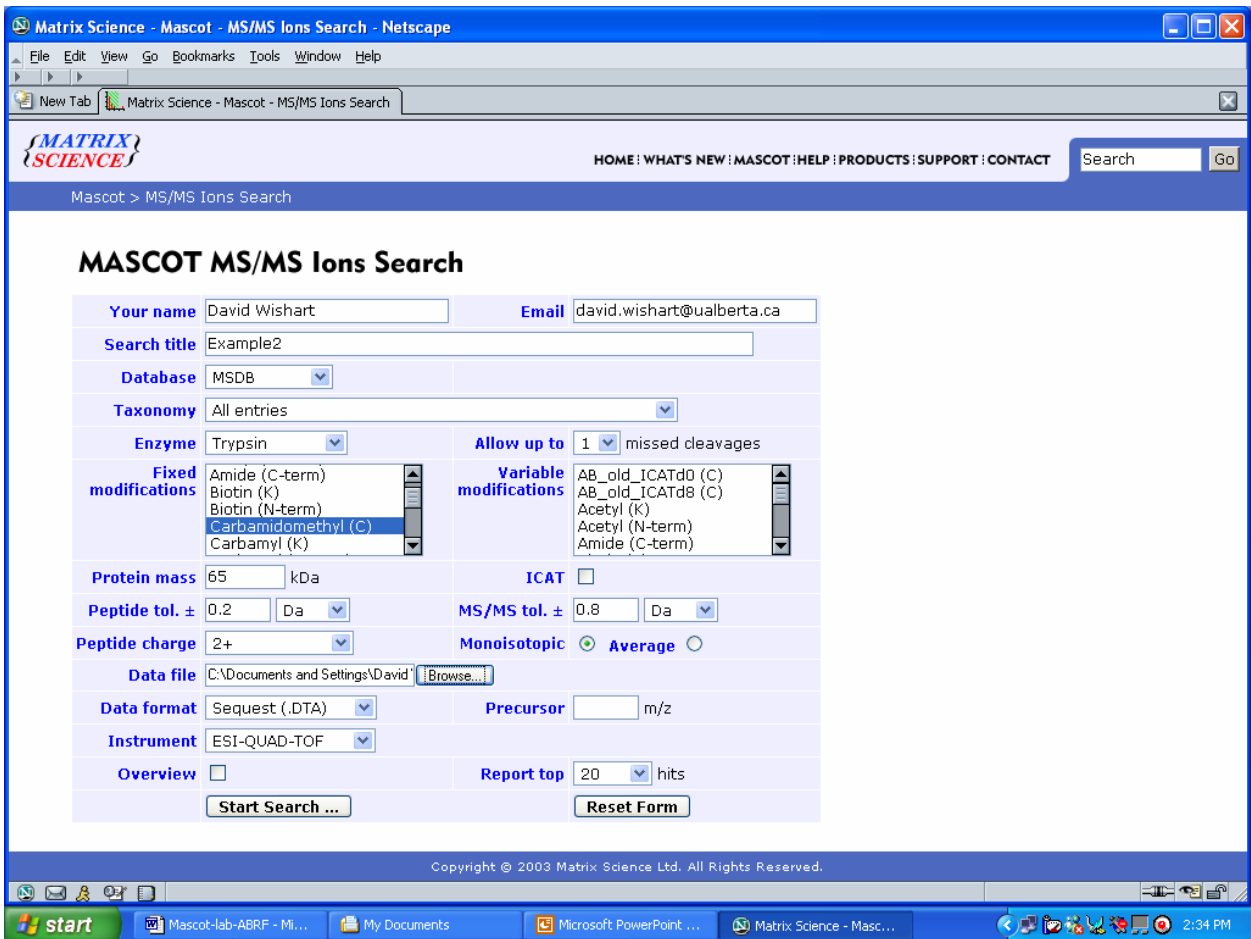
Sort Peptides By Residue Number Increasing Mass Decreasing Mass

Start	End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
37	45	998.60	997.59	997.57	0.03	1	RQLANTPAK
51	66	1805.90	1804.89	1804.89	0.01	0	EVARPNNYAGALYDPR
101	117	884.68	883.68	883.68	0.00	0	DNNYELKVVV

Example 2: MS-MS Fragment Ion Analysis.

Human nuclei sample has been submitted to Iodoacetamide and migrated on 1-D SDS gel. One band has been excised at about 65KDa and processed with trypsin. The peptide mixture has been measured in duplicates with MALDI-TOF MS with a mass accuracy below 0.2 Da. The same sample has also been submitted to LC-MS/MS on a Q-TOF MS instrument. All obtained MS/MS spectra have been submitted to peak detection and concatenated into a single file.

1. Open Netscape or Internet Explorer and go to <http://gchelpdesk.ualberta.ca/ABRF2005/> and look for the 'example2.dta' link. Click on the link to view the file. Save the file to your home folder.
2. Go to the Mascot Home and select the **MS/MS Ion Search Form**.
3. Fill out the form as it appears below:



Some of the new fields you will see here are:

MS/MS tol: This is the Error window for MS/MS fragment ion mass values. Leave at 0.8.

ICAT: this restricts the analysis to cysteine containing residues. Leave **unchecked**.

Data Format: Mascot can parse several. Choose **Sequest (DTA)**.

Precursor: The precursor mass from which the fragments are generated. This information is provided in the DTA file, so **leave blank**.

Instrument: the type of instrument used to acquire the MS/MS data. Choose **ESI-QUAD-TOF**.

4. **Start** the search.

Peptide Summary Report (Example2) - Netscape

File Edit View Go Bookmarks Tools Window Help

New Tab Peptide Summary Report (Example2)

Mascot Search Results

User : David Wishart
Email : david.wishart@ualberta.ca
Search title : Example2
MS data file : Lab 2.5 Example 2.dta
Database : MSDB 20041120 (1721490 sequences; 554556810 residues)
Timestamp : 26 Jan 2005 at 21:35:19 GMT

Significant hits:

Q6PK08	NOP5/NOP58 protein (Fragment).- Homo sapiens (Human).
Q6AYI1	Hypothetical protein.- Rattus norvegicus (Rat).
TRPGTR	trypsin (EC 3.4.21.4) precursor - pig (tentative sequence)
Q9BVP2	Nucleostemin, isoform 1.- Homo sapiens (Human).
Q6PL22	L12.- Homo sapiens (Human).
CAA09374	HSA010840 NID: - Homo sapiens
Q8IX96	Multiple hat domains.- Homo sapiens (Human).
Q8N473	Alpha 1 type I collagen, preproprotein.- Homo sapiens (Human).

Probability Based Mowse Score

Ions score is $-10 \cdot \log(P)$, where P is the probability that the observed match is a random event.
 Individual ions scores > 45 indicate identity or extensive homology ($p < 0.05$).
 Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.

Done Wednesday, January 26, 2005

start Mascot-lab-ABRF - Mi... My Documents Microsoft PowerPoint ... Peptide Summary Re... 2:35 PM

Mascot appears by default with a 'Peptide Score'. Scroll down and look at the protein score for the top hit.

Peptide Summary Report (Example 2) - Netscape

Format As: Peptide Summary

Significance threshold p < 0.05 Max. number of hits 20

Standard scoring: MudPIT scoring Ions score cut-off 0 Show sub-sets

Show pop-ups: Suppress pop-ups Sort unassigned Decreasing Score Require bold red

Select All Select None Search Selected Error tolerant

1. [Q6PK08](#) Mass: 59154 Score: 306 Queries matched: 6
 NOP5/NOP58 protein (Fragment).- Homo sapiens (Human).
 Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
<input checked="" type="checkbox"/> 19	613.83	1225.64	1225.60	0.04	0	52	0.014	1	IISDNLTYCK
<input checked="" type="checkbox"/> 26	662.89	1323.76	1323.71	0.04	0	89	2.7e-06	1	HAASTVQILGAEK
<input checked="" type="checkbox"/> 28	664.36	1326.70	1326.66	0.04	0	44	0.079	1	TQLYEYLQNR
<input checked="" type="checkbox"/> 34	792.95	1583.89	1583.83	0.07	0	25	4.4	1	LSELLPEEVEAEVK
<input checked="" type="checkbox"/> 37	814.89	1627.76	1627.70	0.05	0	65	0.00058	1	TYDPSGDSLPTCSK
<input checked="" type="checkbox"/> 50	941.52	1881.03	1880.96	0.07	0	31	1.1	1	LNLSCIHSPVVELMR

Proteins matching the same set of peptides:

[T17299](#) Mass: 50347 Score: 306 Queries matched: 6
 hypothetical protein DKFZp564H2171.1 - human (fragment)

[AAD27610](#) Mass: 60054 Score: 306 Queries matched: 6
 AF123534 NID: - Homo sapiens

[CAE91962](#) Mass: 52769 Score: 306 Queries matched: 6
 AX883409 NID: - Homo sapiens

2. [Q6AY11](#) Mass: 69709 Score: 250 Queries matched: 6

For each protein, the first line contains the accession number (linked to the corresponding protein view, the protein molecular weight, and the total ions score (not including contributions from peptide mass matches), and the number of peptides matched. This is followed by a brief descriptive title, then a table summarising the matched peptides. The table columns contain:

- Hyper linked query number (see below)
- Experimental m/z value
- Experimental m/z transformed to a relative molecular mass
- Calculated relative molecular mass of the matched peptide
- Difference (error) between the experimental and calculated masses
- Number of missed cleavage sites

- The 5% significance level for the ions score
- Ions score. If there are duplicate matches to the same peptide, then only the highest scoring match contributes to the total score. The lower scoring matches are shown in brackets.
- Rank of the ions match (1 to 10, where 1 is the best match)
- Sequence of the peptide in 1-letter code

Mouse over the query '19' for the first hit. A pop up appears. The popup displays the complete list of peptide matches for that query.

Peptide Summary Report (Example2) - Netscape

Format As: Peptide Summary [Help](#)

Significance threshold $p < 0.05$ Max. number of hits: 20

Standard scoring MudPIT scoring Ions score cut-off: 0 Show sub-sets

Show pop-ups Suppress pop-ups Sort unassigned: Decreasing Score Require bold red

Select All Select None Search Selected Error tolerant

1. [Q6PK08](#) Mass: 59154 Score: 306 Queries matched: 6
NOP5/NOP58 protein (Fragment) -- Homo sapiens (Human).
 Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
<input checked="" type="checkbox"/> 19	613.83	1225.64	1225.60	0.04	0	52	0.014	1	IISDNLTYCK
<input checked="" type="checkbox"/> 26	662.89	1323.76	1323.71	0.04	0	89	2.7e-06	1	HAASTVQILGAEK
<input checked="" type="checkbox"/> 2	Top scoring peptide matches to query 19								
<input checked="" type="checkbox"/> 3	Score greater than 31 indicates homology								
<input checked="" type="checkbox"/> 3	Score greater than 46 indicates identity								
<input checked="" type="checkbox"/> 5	Status bar shows all hits for this peptide								

Score	Delta	Hit	Protein	Peptide
51.9	-0.04	1	Q6PK08	IISDNLTYCK
15.7	-0.03			LLSDPADARLR
14.6	-0.02			IINGDVPETLR
14.5	-0.13			ILSGQKVAILGK
13.9	-0.05			LLSDLPLVNVK
13.8	-0.06			LLSDPRGLLDK
12.3	-0.05			LISDPEDPKVK
11.6	0.04			LISDDYTMLR
11.2	-0.13			LLEAIRTLGLK
11.2	-0.07			LISWRIPVDK

2. [Q6AYI1](#) Mass: 69709 Score: 250 Queries matched: 6

http://www.matrixscience.com/cgi/peptide_view.pl?file=.../data/20050126/FktrmiuE.dat&query=19&hit=1&index=Q6PK08&pr=1

start Mascot-lab-ABRF - Mi... My Documents Microsoft PowerPoint ... Peptide Summary Re... 2:37 PM

Click the query '19' to bring up the peptide view.

Mascot Search Results: Peptide View - Netscape

File Edit View Go Bookmarks Tools Window Help

New Tab Mascot Search Results: Peptide View

Mascot Search Results

Peptide View

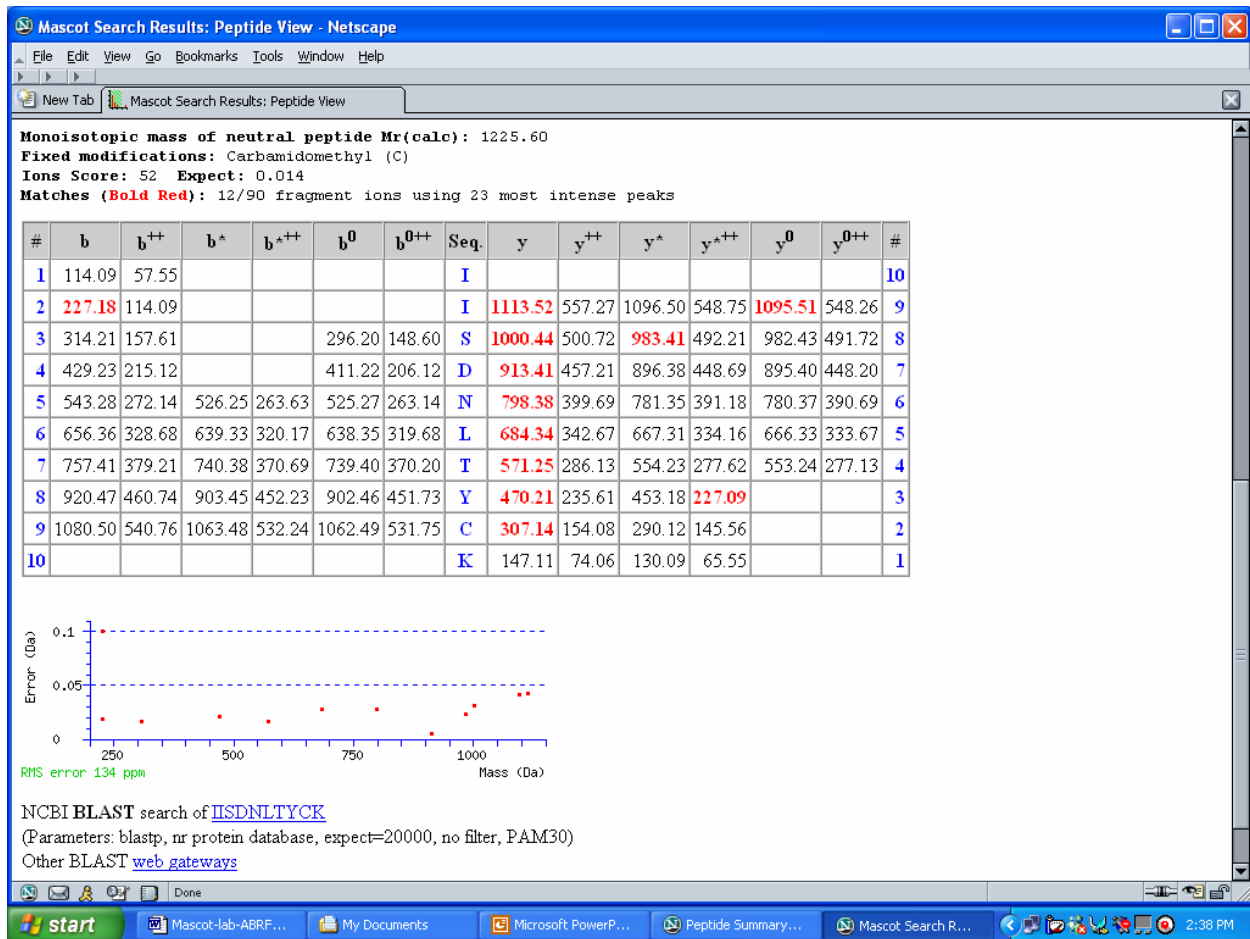
MS/MS Fragmentation of **IISDNLTYCK**
 Found in **Q6PK08**, NOP5/NOP58 protein (Fragment) - Homo sapiens (Human).

Match to Query 19: 1225.644973 from(613.829762,2+)
 From data file Lab_2.5_Example_2.dta

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

Monoisotopic mass of neutral peptide Mr(calc): 1225.60
 Fixed modifications: Carbamidomethyl (C)

start Mascot-lab-ABRF... My Documents Microsoft PowerP... Peptide Summary... Mascot Search R... Synaptics Pointing Device 2:37 PM



Mascot also provides a 'Proteins Summary Report' for MS/MS data. The protein summary report seeks to answer the question: "which single protein best matches the experimental data?". In the case of a complex mixture of proteins, the hit list in a protein summary report tends to be dominated by those proteins which are well represented in the mixture. Proteins with just one or two peptide matches can be so far down the list they are never seen. In this case, the Peptide Summary Report is more appropriate.

Protein Summary Report (Example2) - Netscape

Accession	Mass	Score	Description
1. Q6PK08	59154	359	NOPS/NOP58 protein (Fragment).- Homo sapiens (Human).
2. AAD27610	60054	358	AF123534 NID: - Homo sapiens
3. T17299	50347	354	hypothetical protein DKFZp564H2171.1 - human (fragment)
4. CAF91962	52769	353	AX883409 NID: - Homo sapiens
5. AAH83986	57044	339	BC083986 NID: - Xenopus laevis
6. Q8C8Y7	60804	336	Mus musculus adult retina cDNA, RIKEN full-length enriched library, clone:A930001M05 product:nu
7. Q6DFW4	60818	336	Nol5 protein.- Mus musculus (Mouse).
8. 070396	53591	331	SIK similar protein.- Mus musculus (Mouse).
9. Q7Z2V5	46481	330	Hypothetical protein DKFZp686J01190 (Fragment).- Homo sapiens (Human).
10. AAC23535	52051	328	AF069782 NID: - Rattus sp.
11. AAF05769	60546	323	AF194371 NID: - Rattus norvegicus
12. JC1087	69618	316	RNA helicase, ATP-dependent - human
13. AAP36310	69731	316	Homo sapiens DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 5 (RNA helicase, 68kDa) (Fragment).-
14. I48385	69790	307	RNA helicase TN22 - mouse
15. Q6P5F8	75093	307	Ddx5 protein (Fragment).- Mus musculus (Mouse).
16. Q6AY11	69709	307	Hypothetical protein.- Rattus norvegicus (Rat).
17. Q8BTS0	69736	307	Mus musculus 2 days neonate thymus thymic cells cDNA, RIKEN full-length enriched library, clon
18. I4VWA	24145	275	trypsin (EC 3.4.21.4), chain A - pig
19. I4NLE	24142	275	trypsin inhibitor, chain E - medicinal leech (fragments)
20. TRPGTR	25078	274	trypsin (EC 3.4.21.4) precursor - pig (tentative sequence)

Results List

1. [Q6PK08](#) **Mass:** 59154 **Score:** 359 **Expect:** 2.2e-30 **Queries matched:** 10

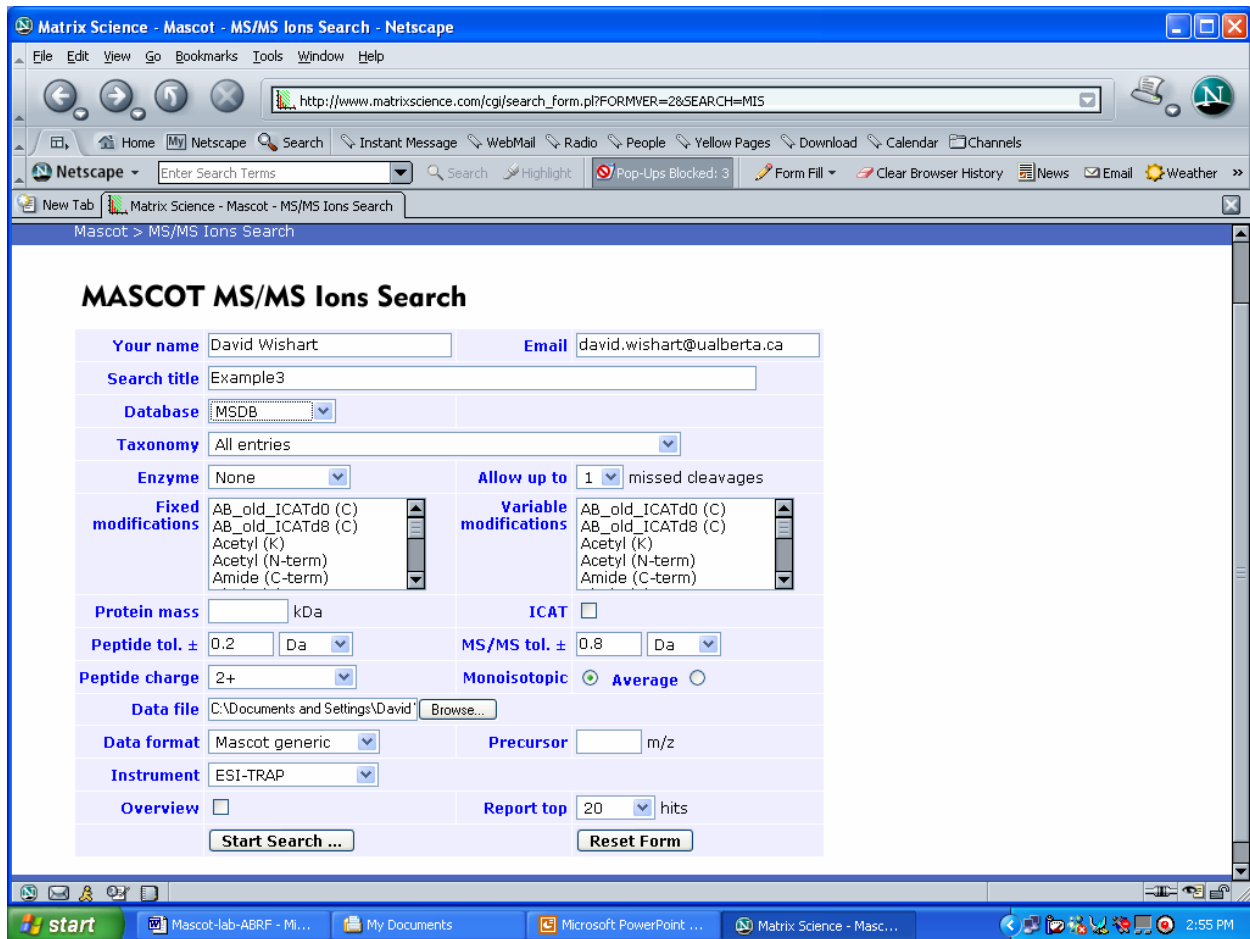
NOPS/NOP58 protein (Fragment).- Homo sapiens (Human).

Observed	Mr(expt)	Mr(calc)	Delta	Start	End	Miss	Ions	Peptide
690.42	689.41	689.41	0.01	405	- 411	1	---	KISGTGK
613.83	1225.64	1225.60	0.04	197	- 206	0	52	IISDNLTYCK
613.87	1225.74	1225.60	0.13	197	- 206	0	---	IISDNLTYCK
662.89	1323.76	1323.71	0.04	311	- 323	0	89	HAASTVQILGAEK
664.36	1326.70	1326.66	0.04	269	- 278	0	44	TQLYEYLQNR
684.89	1367.77	1367.64	0.12	175	- 184	1	---	ELNNYIMRCR
792.95	1583.89	1583.83	0.07	222	- 235	0	25	LSELLPEEVEAEVK
793.93	1585.85	1585.88	-0.02	2	- 15	0	---	LVLFETSVGYAIFK
814.89	1627.76	1627.70	0.05	427	- 441	0	65	TYDPSGGSTLPTCSK

Example 3: MS-MS Fragment Ion Analysis of Neurosensin.

A novel neuropeptide hormone induced by music/sound has been found in vertebrates. It is likely a cleavage product of a larger protein. Preliminary MS and NMR analysis suggests that there are no known or suspected PTMs (post-translational modifications). MS/MS spectra were collected on an ion trap spectrum. Your task is to identify what this hormone is (its sequence, parent protein, whatever).

- Open Netscape or Internet Explorer and go to <http://gchelpdesk.ualberta.ca/ABRF2005/> and look for the 'example3.mgf' link. Click on the link to view the file. Save the file to your home folder.
- Go to the Mascot Home and select the **MS/MS Ion Search Form**.
- Fill out the form as it appears below:



7. Start the search. Use the techniques and tricks you learned in the previous 2 examples to see if you can get a meaningful result. This last problem is more open-ended and we won't guide you through all the steps to get the answer.

Appendix

1. Resources

i) Original Papers

- Perkins, DN, Pappin, DJ, Creasy, DM and Cottrell, JS (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**:3551-3567.

- Pappin, DJC, Hojrup, P and Bleasby AJ, (1993) Rapid identification of proteins by peptide-mass fingerprinting. *Curr. Biol.* 3:327-32.

ii) Web Sites:

- Matrix Science: <http://www.matrixscience.com>

Acknowledgements

Thanks to Dr. Gary Van Domselaar (University of Alberta) for his assistance and insight in helping to prepare these examples