

MASCOT *Integra*

MCP Guidelines without tears

MASCOT : MCP Guidelines without tears

© 2007 Matrix Science



An introduction to publication guidelines

A Review/Perspective in Molecular & Cellular Proteomics mentions the idea of guidelines

Mol Cell Proteomics. 2004 Jan;3(1):1-9.

MCP guidelines first introduced in 2004

Mol Cell Proteomics. 2004 Jun;3(6):531-3.

MASCOT : MCP Guidelines without tears

© 2007 Matrix Science



As any of you who have recently submitted a manuscript for publication will have noticed, many journals have introduced guidelines that cover the reporting of Mass spectrometry data acquisition, processing and search results. The proteomics community is not the only community to publish such guidelines, similar guidelines have been produced for the microarray and microscope imaging communities. One of the first journals in the proteomics community to put forward guidelines was the journal Molecular and Cellular Proteomics.

MCP published a review by Mike Baldwin, with input from the journals editors, on the reporting methods used in proteomics experiments and suggested a number of guidelines for publication.

The MCP editor Steve Carr then chaired a small working group that drafted an initial set of guidelines that were published in 2004.

Publication guidelines

The Paris meeting was attended by representatives from the manufacturing and publishing industries as well as scientists and the guidelines were made available for public comment.

http://www.mcponline.org/misc/ParisReport_Final.shtml

MCP are not the only journal to provide publication guidelines

Guidelines for the next 10 years of proteomics, *Proteomics* 6(1) 4-8 2006

MASCOT : *MCP Guidelines without tears*

© 2007 Matrix Science

MATRIX
SCIENCE

These guidelines were revised and expanded a year later in a larger meeting held in Paris. After public review they were published in 2006 and are currently in effect.

Around the same time the journal *Proteomics* also published guidelines for authors submitting manuscripts to the journal. Many other journals use these guidelines as starting points for their own reviewers. Although we refer to the reports produced by the Mascot Integra data management system in this presentation as meeting the MCP guidelines they also meet the *Proteomics* journals guidelines.

Why are guidelines and standards worth having?

- Proteomics is complex
- Multiple experimental and data processing options
- Ensures that reviewers have sufficient information
- Essential for credibility and reputation
- Encourages use of standard data formats

MASCOT : MCP Guidelines without tears

© 2007 Matrix Science

MATRIX
SCIENCE

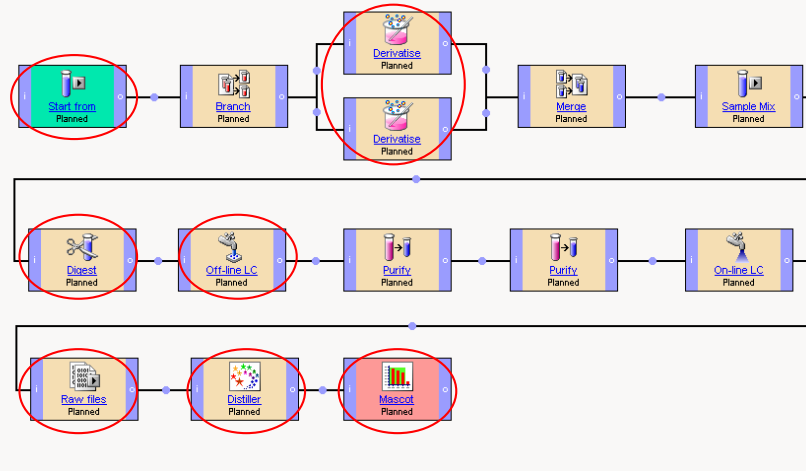
Proteomics is complex, with many options available during both the experimental stages and data analysis. The parameters used need to be clearly documented. For a reviewer or reader to fully comprehend the results certain information needs to be presented. Guidelines help by defining what should be included.

The community wants to keep and improve its scientific reputation and credibility by encouraging the publication of high quality papers. Papers that place great significance on poorly presented or incomplete data can ruin the reputation of the community. Guidelines can help everyone design valid experiments and report high quality results.

Reporting standards can be more stringent than guidelines in that they normally define a vocabulary and format for the information. However a standard might not require all the information that guidelines recommend.

Now on to modeling an experiment in Integra and capturing the information that is to be reported.

MS/MS Example: ICAT wash through



MASCOT : MCP Guidelines without tears

© 2007 Matrix Science

MATRIX
SCIENCE

Here we have modelled an MS/MS experiment in Mascot Integra which supports the ICAT procedure, but where we are going to analyse the peptides that did not bind to the avidin column. Protein extracts from G1 and M-phase arrested *S.cerevisiae* cells were labelled with the ICAT light and heavy reagents respectively. The labelled extracts are then mixed, digested, separated into 6 fractions and then purified before being subjected to MS/MS analysis. Here we're just picking up the existing raw data files, then we carryout peak detection using Mascot Distiller before searching SwissProt release 51.6 with Mascot 2.2.

Experimental data capture

EXPERIMENTTASKS
EXP-070500389-1907 **Digest** In-progress

Digestion Details Variant:1 Instance:1

Parameter	Type	Rep	Entered Value	Unit
Protocol Id	Standard	1		
Digestion buffer	Standard	1	25 mM Ammonium B	
Digestion buffer vol	Standard	1	50	
Buffer volume units	Standard	1		µl
Incubation tempo	Standard	1	37	
Incubation Time	Standard	1	4	
Incubation time unit	Standard	1		hours
Storage type	Standard	1		
Instrument Id	Standard	1		
Stop buffer used	Standard	1		
Stop buffer volume	Standard	1		
Stop buffer units	Standard	1		

Mascot Enzyme Variant:1 Instance:1

Parameter	Type	Rep	Entered Value	Unit
Enzyme id	Standard	1	Trypsin	
Enzyme Volume used	Standard	1	2.5	
Enzyme volume units	Standard	1	µg	

Sample volume used Variant:1 Instance:1

Parameter	Type	Rep	Entered Value	Unit
Sample volume used	Standard	1	100	
Sample volume units	Standard	1		µl

MASCOT : MCP Guidelines without tears © 2007 Matrix Science **MATRIX SCIENCE**

While you run an experiment in Mascot Integra you can capture information about how each experimental step was carried out. In this example we're showing the data capture page for the 'Digest' task and capturing details about how the protein extracts were digested with trypsin. Most of the fields are optional. However, the more information we capture about the experiment as we run it, the more information we can automatically include in our MCP publication report when we export it from Mascot Integra.

Raw data reduction

Distiller(EXP-070500389-1913) Define sample raw data reduction

Multi-sample files:

Merge all samples into a single search

Select Range (multi-scan files):

Start End Units

Peak List Format:

MGF
Comprehensive
mzData
 Save

Output PMF Masses as: Output MS/MS Fragments as:

m/z MH+ Mr m/z MH+ Mr

MASCOT : MCP Guidelines without tears

© 2007 Matrix Science

MATRIX
SCIENCE

During the raw data reduction step using Mascot Distiller we're going to export the peaklists as mzData files. mzData is an information rich peaklist file format specified by the Proteomics Standards Initiative workgroup (PSI) of HUPO. If you're going to use Mascot Integra (or any other system) to generate an MCP compliant report then I would recommend that you use this data format. There are several reasons for this:

1. It is a requirement of the mzData format that the software (and version number) used to generate the peaklist is included in the file. Also some of the peak detection parameters are often included. We need this information to generate the MCP report and so using the mzData format removes the requirement to prompt the user for the information.
2. MCP guideline 8 encourages you to include the MS/MS spectra mentioned in the paper as supplementary material. mzData is an acceptable format for this, and since it includes details of how the data reduction was carried out seems to be a sensible choice.
3. If you choose to submit your experiment to PRIDE, mzData is also the required format for submitting peaklists, so generating the peaklists as mzData files here could save you time later on.

Protein hit approval

Peptide summary report

Selected hit: **IMP4_YEAST** Importin beta-4 subunit (Karyopherin beta-4 subunit) (Karyopherin-123) (Ran-binding protein YRB4) - Saccharomyces cerevisiae (Bakers yeast)

Check to approve protein and peptide matches.

Comments:

IMP4_YEAST Mass: 122524 Total Score: 283 Queries Matched: 31 emPAI: 0.45
Importin beta-4 subunit (Karyopherin beta-4 subunit) (Karyopherin-123) (Ran-binding protein YRB4) - Saccharomyces cerevisiae (Baker's yeast)

Match?	Query	Observed	Mr(exp)	Mr(Calc)	ppm	Miss	Score	Expect	Rank	Peptide
<input type="checkbox"/>	5	357.73	713.44	713.47	-35	0	(30)	0.0186	1	K.VIELLK.Y
<input checked="" type="checkbox"/>	6	357.73	713.45	713.47	-33	0	45	0.0007	1	K.VIELLK.Y
<input checked="" type="checkbox"/>	9	359.22	716.42	716.44	-30	0	49	0.0007	1	K.SVLLASK.Y
<input checked="" type="checkbox"/>	116	422.23	842.44	842.46	-23	0	53	0.0001	1	K.QLAGVEAR.K
<input type="checkbox"/>	239	474.74	947.46	947.46	-3	0	27	0.0474	1	K.VEPESYPK.G
<input type="checkbox"/>	254	480.78	959.54	959.57	-29	0	28	0.0526	1	K.TILPEIFK.T
<input type="checkbox"/>	255	480.78	959.55	959.57	-21	0	(23)	0.1912	1	K.TILPEIFK.T
<input type="checkbox"/>	278	497.26	992.50	992.50	-5	0	28	0.0407	1	K.YLDPIMNK.L
<input checked="" type="checkbox"/>	319	520.74	1039.47	1039.46	1	0	40	0.0011	1	K.QFQTEENK.H
<input type="checkbox"/>	380	556.33	1110.65	1110.66	-14	0	(29)	0.0183	1	K.LGPETTYAALK.V
<input checked="" type="checkbox"/>	391	556.33	1110.65	1110.66	-12	0	46	0.0004	1	K.LGPETTYAALK.V
<input checked="" type="checkbox"/>	508	629.81	1257.61	1257.62	-8	0	43	0.0013	1	K.FTVNTGISYEK.E
<input type="checkbox"/>	518	634.35	1266.68	1266.72	-34	0	11	2.2947	1	R.IIEIFSAVFTK.E
<input type="checkbox"/>	519	634.35	1266.68	1266.72	-31	0	(0)	29.366	3	R.IIEIFSAVFTK.E
<input type="checkbox"/>	533	642.32	1282.63	1282.66	-17	0	11	1.8589	1	R.ESGYAFIANLAK.V
<input type="checkbox"/>	535	642.81	1283.61	1283.66	-32	0	(0)	20.2943	1	K.VYGENFAPFLK.T
<input checked="" type="checkbox"/>	537	642.82	1283.63	1283.66	-20	0	52	0.0001	1	K.VYGENFAPFLK.T
<input checked="" type="checkbox"/>	559	661.34	1320.67	1320.69	-20	0	57	6.485E-5	1	K.ALYELLSAADQK.A

MASCOT : MCP Guidelines without tears

© 2007 Matrix Science



Once the searches have been completed, for the MS/MS MCP report Mascot Integra only exports protein hits which have been approved in Mascot Integra. This is done from reports similar to the standard Mascot reports, but generated from within Integra. Integra has a number of features which help in this process. Here we've set the report probability threshold to 0.01, and got the system to flag up any peptides with significant ions scores. You can also write flexible filters which will filter out protein hits which do not match our approval criteria. For example, we could write a filter which will only show protein hits with at least 2 peptides significant at a p value of 0.05. We are also currently working on fully automated approvals so that you will be able to specify the filter before you carry out the search, and the system will automatically approve protein hits when the search is completed and results are imported into Integra.

Generating the report

The screenshot displays the Mascot software interface for generating an MCP publication report. The top navigation bar includes buttons for 'Return', 'Add', 'Edit', 'View', 'Delete', 'Edit Exp. Plan', 'Start Exp.', and 'Review Exp.'. The 'Publication Report' button is circled in red. Below this, the 'Export publication report' form is shown. The 'MCP MS/MS report parameters' section includes fields for 'Peptide probability threshold' (0.01), 'Calibration method', 'Exclusion of contaminant ions', 'Resolution', 'Mascot Distiller version' (2.1.0.0), and 'Sequence database release' (51.6), which is circled in red. The 'Export sample preparation details' checkbox is checked. At the bottom of the form, the 'Export report' button is circled in red. The URL 'MCP report for experiment EXP-070500389' is visible below the form. The footer contains the text 'MASCOT : MCP Guidelines without tears', '© 2007 Matrix Science', and the 'MATRIX SCIENCE' logo.

Once you've approved the protein hits, generating the MCP publication report couldn't be easier. From the experiment list in Mascot Integra, select the experiment you interested in and then click on the 'Publication Report' button. On the page that opens select the MCP report option and click on 'Export report'. However, the system then needs to capture some additional information. For example, we need the Sequence database release version (or details of the sources of protein sequences if you searched custom database). Other details such as the peak detection software version have been automatically filled in from the details taken from the mzData peaklist files. Once you've filled in the additional values, click on the 'Export report' button again to generate and download the report which consists of the main Excel report and some supporting html files.

Molecular Cellular Proteomics publication guidelines

Guideline 1: Supporting information

- Method used to generate the peaklist
- Database search strategy
- Sequence database used.

MASCOT : MCP Guidelines without tears

© 2007 Matrix Science



OK, now lets review the Molecular Cellular Proteomics publication guidelines to see what information we have to provide, and then look at the report from Mascot Integra to see how we've supported that requirement.

The Molecular Cellular Proteomics publication guideline 1 specifies what supporting information should be reported with data.

The method and/or program (including version number) used to create the "peak list" from the raw data and the parameters used in the creation of this peak list

The name and version of the program(s) used for database searching and the values of search parameters

And the name and version of the database used, or details of the protein sequence sources for a custom database.

Guideline 1: Peak detection software

The screenshot displays an Excel spreadsheet titled 'MCP_Report_Peak_List...'. The spreadsheet is organized into several sections:

- Sample preparation conditions:** Rows 1-18 detailing the experimental workflow, including sample collection, digestion, and purification steps.
- Peak picking parameters:** Rows 19-35, highlighted in red, listing various parameters used for peak detection, such as 'Database search conditions set 1', 'Search engine', 'Database', 'Taxonomy', and 'Protein hit assignment criteria'.
- Database search parameters:** Rows 36-41, detailing search engine settings like 'Search engine', 'Database', 'Taxonomy', 'Database Size', and 'Taxonomy'.
- Search results:** Rows 42-54, showing a list of identified proteins with columns for 'Search ID', 'Search title', 'Score', and 'P-value'.

At the bottom of the spreadsheet, there is a footer with the text: 'MASCOT : MCP Guidelines without tears © 2007 Matrix Science' and the Matrix Science logo.

The program and version number used to create the peaklist from the raw data are shown here on the first tab in the Excel report. We must also supply the parameters used in the creation of the peak list. This information has been put in a separate tab - 'Peak detection parameters'. Here we can clearly see the advantage of having used the mzData peaklist format in conjunction with Mascot Distiller as we have all the required information in a standard, controlled, format. Any reviewer would have all the information required to easily recreate the peak detection settings in Mascot Distiller.

Guideline 1: Database search parameters

The screenshot displays the Mascot software interface with several key sections highlighted:

- Sample preparation conditions:**
 - 1. S.cerevisiae ICAT wash through experiment
 - 2. Sample preparation conditions
 - 3. Proteolytic digestion conditions (Trypsin): 100 µl of each sample were diluted using 50 µl of 25 mM Ammonium Bicarb and subjected to proteolytic digestion using 2.5 µg of Trypsin at 37 degrees celcius for 4 hours.
 - 7. Chemical denaturation step 1: 10 µg of each sample were diluted using 100 µl of ICAT - light (Applied Biosystems, Foster City, CA) and incubated at 21 degrees celcius for 1 hours.
 - 9. Chemical denaturation step 2: 10 µg of each sample were diluted using 50 µl of ICAT - heavy (Applied Biosystems, Foster City, CA) and incubated at 21 degrees celcius for 1 hours.
 - 11. Purification step 1: Samples were purified by Elution. Samples were diluted using 10 µl of Biotin elution buffer.
 - 13. Purification step 2: Samples were purified by Elution. Samples were diluted using 10 µl of Ziptip elution buffer. The purified samples were then washed 3 times in 100 µl of dH2O.
 - 15. Mass spec analysis: MS analyses were carried out using a API QStar Pulsar i.
- Database search parameters:**
 - 22. Database search conditions set 1
 - 23. Search engine: Mascot 2.2.1
 - 24. Database: SwissProt 51.6
 - 25. Database Size: 217964
 - 26. Taxonomy: Fungi
 - 27. Taxonomy Includ: 4751
 - 28. Database Size after Taxonomy: 16473
 - 29. Peptide Mass Accuracy: 50 ppm
 - 30. MS/MS Mass Accuracy: 0.3 Da
 - 31. Maximum missed cleavages: 2
 - 32. Fixed modifications: Acetyl (Protein Nterm)
 - 33. Variable modifications: Tyrosin
 - 34. Enzyme: NA
 - 35. Resolution: NA
 - 36. Calibration: NA
 - 37. Exclusion of contaminant ions: NA
- MCP MS/MS report parameters:**
 - Peptide probability threshold*: 0.01
 - Calibration method*: [Dropdown]
 - Exclusion of contaminant ions*: [Dropdown]
 - Resolution*: [Dropdown]
 - Mascot Distiller version*: 2.1.0.0
 - Sequence database release*: 51.6
 - Export sample preparation details:
 - Protein approval method(s): Proteins must have at least one bold significant peptide match with a p value < 0.01. Only significant peptide matches were approved.
- Protein Hit assignment criteria:** Proteins must have at least one bold significant peptide match with a p value < 0.01. Only significant peptide matches were assessed.
- Search Results Table:**

SearchID	Search title	Source file	Protein Accession No	Protein Description	Mascot Score	Number matched peaks	Sequence Coverage	Number unique p
43	ms-15052007-00003	<Biomap> P1 25-26_5001	Vgapevic_dsweltGDP1_ASH2CO	GTP-binding nuclear protein	67.8022696	2	5.14	2
44	ms-15052007-00003	<Biomap> P1 25-26_5001	Vgapevic_dsweltRMS1_YEAST	Impromin subunit alpha (Ipa)	64.13	2	2.77	1
45	ms-15052007-00003	<Biomap> P1 25-26_5001	Vgapevic_dsweltRSM1_YEAST	40S ribosomal protein S1E	60.79	1	6.25	1
46	ms-15052007-00003	<Biomap> P1 25-26_5001	Vgapevic_dsweltEF1A1_YEAST	Elongation factor 1 alpha (I)	54.4402096	3	5.08	3
47	ms-15052007-00003	<Biomap> P1 25-26_5001	Vgapevic_dsweltDYWC_YEAST	Lysyl-tRNA synthetase, cy	52.6372655	3	4.23	2
48	ms-15052007-00003	<Biomap> P1 25-26_5001	Vgapevic_dsweltKPO1_YEAST	Exonin-1 (Chromosome I)	49.2902096	2	1.48	2
49	ms-15052007-00004	<Biomap> P2 26_5-26002	Vgapevic_dsweltKWC2_YEAST	Endonase 2 (EC 4.2.1.11) (E)	64.63	2	4.58	2
50	ms-15052007-00004	<Biomap> P2 26_5-26002	Vgapevic_dsweltRMB1_YEAST	Impromin beta-4 subunit (Ic)	237.1729996	13	11.32	13
51	ms-15052007-00004	<Biomap> P2 26_5-26002	Vgapevic_dsweltGDP1_ASH2CO	GTP-binding nuclear protein	102.0905999	4	9.81	4
52	ms-15052007-00004	<Biomap> P2 26_5-26002	Vgapevic_dsweltEF1A1_YEAST	Elongation factor 1 alpha (I)	134.7014998	7	10.48	7
53	ms-15052007-00004	<Biomap> P2 26_5-26002	Vgapevic_dsweltKPO1_YEAST	Exonin-1 (Chromosome I)	115.1391664	7	6.19	7
54	ms-15052007-00004	<Biomap> P2 26_5-26002	Vgapevic_dsweltSMX1_YEAST	Impromin beta-SMX1 (Iaryc)	116.1026999	6	5.81	6

MASCOT : MCP Guidelines without tears

© 2007 Matrix Science



Guideline 1 also states that the name and version of the software used for database searching, and the names and values of search parameters, be stated. In addition, it requires that the name and version of the sequence database used be stated and this is included in the Database search parameters. It also requires that the methods used to interpret the MS/MS data be stated. This information was captured as part of the additional information entered at the start of the exporting procedure.

Guideline 1: False-positive rates

File	SwissProt	Decoy	%
mss-15052007-00003	18.00	0.00	0.00
Peptide matches above identity threshold	29.00	0.00	0.00
Peptide matches above homology or identity threshold			
mss-15052007-00004	152.00	4.00	2.63
Peptide matches above identity threshold	187.00	4.00	2.14
Peptide matches above homology or identity threshold			
mss-15052007-00005	52.00	2.00	3.85
Peptide matches above identity threshold	67.00	3.00	4.48
Peptide matches above homology or identity threshold			
mss-15052007-00006	44.00	1.00	2.27
Peptide matches above identity threshold	54.00	1.00	1.85
Peptide matches above homology or identity threshold			
mss-15052007-00007	138.00	0.00	0.00
Peptide matches above identity threshold	161.00	1.00	0.62
Peptide matches above homology or identity threshold			
mss-15052007-00008	66.00	3.00	4.55
Peptide matches above identity threshold	82.00	4.00	4.88
Peptide matches above homology or identity threshold			

MASCOT : MCP Guidelines without tears

© 2007 Matrix Science



Finally Guideline 1 also requires that for large scale experiments you should report any additional statistical analyses that indicate or establish a measure of identification certainty, or allow a determination of false-positive rate. Mascot 2.2 introduced the option to carryout an automatic search against a decoy database. For this experiment, we used this option and so we can export this information as part of the report.

Molecular Cellular Proteomics publication guidelines

Guideline 2: Information for each protein sequence identified should specify the following:

- Accession number and database source;
- Score(s) and any associated statistical information obtained for searches conducted;
- Sequence coverage
- Total number of peptides assigned to the protein.

MASCOT : MCP Guidelines without tears

© 2007 Matrix Science



The second guideline specifies the protein information that should be reported. This includes details about the protein accession and database source

The protein score with any statistics

The percentage sequence coverage of the hit

The total number of peptide sequences assigned to the hit

Guideline 2: Protein identifications

Search ID#	Source file	Protein Accession ID	Protein Description	Mascot Score	Number matched peaks	Sequence Coverage	Number unique peptides	Protein Mass	pI
42	01-25-26_5001	SPY1_ASHGO	GTP-binding nuclear protein importin subunit alpha	67.68	2	2.77	1	24282.48	6.44
43	01-25-26_5001	MA1_YEAST	Importin subunit alpha (p4)	64.13	2	2.77	1	60403.67	4.8
44	01-25-26_5001	RS1A_YEAST	40S ribosomal protein S11	60.79	1	6.29	1	15907.33	9.61
45	01-25-26_5001	EF1A_YEAST	Elongation factor 1-alpha	54.64	3	5.68	3	50001.2	9.34
46	01-25-26_5001	SYK2_YEAST	Lysyl-tRNA synthetase, e	62.64	3	4.23	3	67915.22	5.78
47	01-25-26_5001	IPY1_YEAST	Exportin-1 (Chromosomes 4)	49.29	2	1.68	2	124623.23	5.32
48	01-25-26_5001	SYK1_YEAST	Enolase 2 (EC:4.2.1.11) (c	64.62	2	4.58	2	46951.91	5.67
49	01-25-26_5001	MEI1_YEAST	Importin beta-4 subunit (K	237.17	13	11.32	10	122524.69	4.54
50	01-25-26_5001	SPY1_ASHGO	GTP-binding nuclear prote	192.69	4	9.81	4	24282.48	6.44
51	01-25-26_5001	EF1A_YEAST	Elongation factor 1-alpha	134.70	7	10.48	6	50001.2	9.34
52	01-25-26_5001	IPY1_YEAST	Exportin-1 (Chromosomes 4)	116.14	7	6.16	7	124623.23	5.32
53	01-25-26_5001	SMY1_YEAST	Importin beta-3 subunit (K	115.62	6	5.61	6	108235.58	4.61
54	01-25-26_5001	MEO1_YEAST	Importin beta-3 subunit (K	111.97	3	2.82	3	120563.9	4.61
55	01-25-26_5001	SYK2_YEAST	Lysyl-tRNA synthetase, e	99.98	8	5.08	8	67915.22	5.78
56	01-25-26_5001	VPS1_YEAST	Vacuolar protein sorting-a	96.33	2	3.13	2	78687.66	7.69
57	01-25-26_5001	KUP2_YEAST	Nucleoporin NUP2 (Nucle	84.92	4	4.72	4	77623.66	6.91
58	01-25-26_5001	HSY1_YEAST	Heat shock protein SSB1	84.16	2	3.59	2	66561.56	5.32
59	01-25-26_5001	RS1A_YEAST	40S ribosomal protein S11	81.54	2	11.11	2	15907.33	9.61
60	01-25-26_5001	RL2_YEAST	60S ribosomal protein L12	81.07	2	1.52	2	17811.65	9.43
61	01-25-26_5001	NAT1_YEAST	Ran GTPase activating pro	73.87	2	4.67	2	45787.41	4.83
62	01-25-26_5001	MGI1_YEAST	Phosphoglycerate mutase	73.14	1	4.45	1	27591.58	8.01
63	01-25-26_5001	RSY1_YEAST	40S ribosomal protein S29	71.82	1	9.78	1	9739.08	9.76
64	01-25-26_5001	RL1_YEAST	60S ribosomal protein L18	70.37	1	5.29	1	21681.14	11.4
65	01-25-26_5001	RSY1_YEAST	40S ribosomal protein S29	70.28	1	5.41	1	22818.01	9.84
66	01-25-26_5001	RL21A_SCHPO	60S ribosomal protein L21	64.19	1	5.00	1	18368.74	10.2
67	01-25-26_5001	RSY1_YEAST	40S ribosomal protein S11	64.10	1	7.30	1	14527.71	10.7
68	01-25-26_5001	RL2_YEAST	60S ribosomal protein L12	62.64	1	4.31	1	27217.76	11
69	01-25-26_5001	DBP1_ASHGO	Histone H2B-1 + AbhyA g	61.81	3	13.64	2	14275.83	10.1
70	01-25-26_5001	DBP2_YEAST	Oxythiopeptidase	60.88	2	5.63	2	51785.96	7.6
71	01-25-26_5001	EF2_YEAST	Elongation factor 2 (EF-2)	60.48	2	2.28	2	93230.23	6.92
72	01-25-26_5001	MA1_YEAST	Importin subunit alpha (p4)	57.61	2	2.77	2	60403.67	4.8
73	01-25-26_5001	IPY1_YEAST	Exportin-1 (Chromosomes 4)	56.84	3	6.26	3	145134.74	5.66
74	01-25-26_5001	IPY1_YEAST	Exportin-1 (Chromosomes 4)	54.24	1	6.87	1	13816.63	10.6
75	01-25-26_5001	RSY1_YEAST	40S ribosomal protein S29	53.06	3	6.36	3	14618.96	9.94
76	01-25-26_5001	RL18B_YEAST	60S ribosomal protein L18	52.61	2	18.00	2	11128.31	11.6
77	01-25-26_5001	RSY1_YEAST	40S ribosomal protein S29	51.07	1	9.99	1	13898.55	9.52
78	01-25-26_5001	RSY1_YEAST	40S ribosomal protein S29	49.53	2	2.95	2	5475.76	4.52
79	01-25-26_5001	MA1_YEAST	Importin subunit alpha (p4)	49.09	1	7.62	1	12731.51	7.3
80	01-25-26_5001	MA1_YEAST	Importin subunit alpha (p4)	48.68	1	1.55	1	4907.93	7.12
81	01-25-26_5001	RSY1_YEAST	40S ribosomal protein S29	47.25	2	5.28	2	40029.69	6.23
82	01-25-26_5001	RL13A_YEAST	60S ribosomal protein L13	46.53	2	10.55	2	22540.43	11.2
83	01-25-26_5001	RSY1_YEAST	40S ribosomal protein S29	46.46	1	7.08	1	14225.07	10.16
84	01-25-26_5001	RSY1_YEAST	40S ribosomal protein S29	46.46	1	6.57	2	44668.99	6.92
85	01-25-26_5001	MA1_YEAST	Importin subunit alpha (p4)	42.41	3	2.48	3	119996.66	4.97
86	01-25-26_5001	IPY1_YEAST	Exportin-1 (Chromosomes 4)	42.41	2	2.97	2	46951.91	5.67
87	01-25-26_5001	MEO1_YEAST	Importin beta-3 subunit (K	57.45	1	0.78	1	119996.66	4.97
88	01-25-26_5001	SPY1_ASHGO	GTP-binding nuclear prote	50.75	1	2.42	1	35322.39	6.4
89	01-25-26_5001	MA1_YEAST	Importin subunit alpha (p4)	45.38	1	1.28	1	60403.67	4.8
90	01-25-26_5001	RSY1_YEAST	40S ribosomal protein S11	41.32	1	4.86	1	15907.33	9.61
91	01-25-26_5001	NEF2_YEAST	ACE2-binding factor 2, mod	38.87	1	4.37	1	21548.4	9.68
92	01-25-26_5001	RSY1_YEAST	40S ribosomal protein S11	38.20	1	3.05	1	22479.14	9.10
93	01-25-26_5001	RSY1_YEAST	40S ribosomal protein S11	37.40	1	1.10	1	60027.52	9.88

MASCOT : MCP Guidelines without tears

© 2007 Matrix Science



On the first tab of the exported report, below the search parameters, we provide the protein hit information. Here we have the accession number. The database source is provided as part of the search parameters. The protein scores are recalculated to include only the peptide matches that we approved. The percentage sequence coverage, and the total number of unique peptide sequences assigned to the protein, where multiple charge state observations for the same peptide sequence are counted as one (again this is taken only from the peptide matches that were approved).

Guideline 2: Peptides

QueryID	Rank	Observed charge	Observed modification	Protein name	Molecular weight	pI	Ions score	E-value	Start residue	End residue	Missed cleavages	Peptide sequence	Variable modifications
1	1	2		binding nuclear protein GSP1R	1014.571	10.4	57	8E-04	10	20	0	KLVVGGGGTGT	
2	2	2		Abbya gonypil	1014.571	10.4	41	2.9E-3	10	20	0	KLVVGGGGTGT	
3	3	2		Abbya gonypil	1014.571	10.4	57	8E-04	10	20	0	KLVVGGGGTGT	
4	4	2		Abbya gonypil	1014.571	10.4	41	2.9E-3	10	20	0	KLVVGGGGTGT	
5	5	2		Abbya gonypil	1014.571	10.4	57	8E-04	10	20	0	KLVVGGGGTGT	
6	6	2		Abbya gonypil	1014.571	10.4	41	2.9E-3	10	20	0	KLVVGGGGTGT	
7	7	2		Abbya gonypil	1014.571	10.4	57	8E-04	10	20	0	KLVVGGGGTGT	
8	8	2		Abbya gonypil	1014.571	10.4	41	2.9E-3	10	20	0	KLVVGGGGTGT	
9	9	2		Abbya gonypil	1014.571	10.4	57	8E-04	10	20	0	KLVVGGGGTGT	
10	10	2		Abbya gonypil	1014.571	10.4	41	2.9E-3	10	20	0	KLVVGGGGTGT	
11	11	2		Abbya gonypil	1014.571	10.4	57	8E-04	10	20	0	KLVVGGGGTGT	
12	12	2		Abbya gonypil	1014.571	10.4	41	2.9E-3	10	20	0	KLVVGGGGTGT	
13	13	2		Abbya gonypil	1014.571	10.4	57	8E-04	10	20	0	KLVVGGGGTGT	
14	14	2		Abbya gonypil	1014.571	10.4	41	2.9E-3	10	20	0	KLVVGGGGTGT	
15	15	2		Abbya gonypil	1014.571	10.4	57	8E-04	10	20	0	KLVVGGGGTGT	
16	16	2		Abbya gonypil	1014.571	10.4	41	2.9E-3	10	20	0	KLVVGGGGTGT	
17	17	2		Abbya gonypil	1014.571	10.4	57	8E-04	10	20	0	KLVVGGGGTGT	
18	18	2		Abbya gonypil	1014.571	10.4	41	2.9E-3	10	20	0	KLVVGGGGTGT	
19	19	2		Abbya gonypil	1014.571	10.4	57	8E-04	10	20	0	KLVVGGGGTGT	
20	20	2		Abbya gonypil	1014.571	10.4	41	2.9E-3	10	20	0	KLVVGGGGTGT	
21	21	2		Abbya gonypil	1014.571	10.4	57	8E-04	10	20	0	KLVVGGGGTGT	
22	22	2		Abbya gonypil	1014.571	10.4	41	2.9E-3	10	20	0	KLVVGGGGTGT	
23	23	2		Abbya gonypil	1014.571	10.4	57	8E-04	10	20	0	KLVVGGGGTGT	
24	24	2		Abbya gonypil	1014.571	10.4	41	2.9E-3	10	20	0	KLVVGGGGTGT	
25	25	2		Abbya gonypil	1014.571	10.4	57	8E-04	10	20	0	KLVVGGGGTGT	
26	26	2		Abbya gonypil	1014.571	10.4	41	2.9E-3	10	20	0	KLVVGGGGTGT	
27	27	2		Abbya gonypil	1014.571	10.4	57	8E-04	10	20	0	KLVVGGGGTGT	
28	28	2		Abbya gonypil	1014.571	10.4	41	2.9E-3	10	20	0	KLVVGGGGTGT	
29	29	2		Abbya gonypil	1014.571	10.4	57	8E-04	10	20	0	KLVVGGGGTGT	
30	30	2		Abbya gonypil	1014.571	10.4	41	2.9E-3	10	20	0	KLVVGGGGTGT	
31	31	2		Abbya gonypil	1014.571	10.4	57	8E-04	10	20	0	KLVVGGGGTGT	
32	32	2		Abbya gonypil	1014.571	10.4	41	2.9E-3	10	20	0	KLVVGGGGTGT	
33	33	2		Abbya gonypil	1014.571	10.4	57	8E-04	10	20	0	KLVVGGGGTGT	
34	34	2		Abbya gonypil	1014.571	10.4	41	2.9E-3	10	20	0	KLVVGGGGTGT	
35	35	2		Abbya gonypil	1014.571	10.4	57	8E-04	10	20	0	KLVVGGGGTGT	
36	36	2		Abbya gonypil	1014.571	10.4	41	2.9E-3	10	20	0	KLVVGGGGTGT	
37	37	2		Abbya gonypil	1014.571	10.4	57	8E-04	10	20	0	KLVVGGGGTGT	
38	38	2		Abbya gonypil	1014.571	10.4	41	2.9E-3	10	20	0	KLVVGGGGTGT	
39	39	2		Abbya gonypil	1014.571	10.4	57	8E-04	10	20	0	KLVVGGGGTGT	
40	40	2		Abbya gonypil	1014.571	10.4	41	2.9E-3	10	20	0	KLVVGGGGTGT	
41	41	2		Abbya gonypil	1014.571	10.4	57	8E-04	10	20	0	KLVVGGGGTGT	
42	42	2		Abbya gonypil	1014.571	10.4	41	2.9E-3	10	20	0	KLVVGGGGTGT	
43	43	2		Abbya gonypil	1014.571	10.4	57	8E-04	10	20	0	KLVVGGGGTGT	
44	44	2		Abbya gonypil	1014.571	10.4	41	2.9E-3	10	20	0	KLVVGGGGTGT	
45	45	2		Abbya gonypil	1014.571	10.4	57	8E-04	10	20	0	KLVVGGGGTGT	
46	46	2		Abbya gonypil	1014.571	10.4	41	2.9E-3	10	20	0	KLVVGGGGTGT	
47	47	2		Abbya gonypil	1014.571	10.4	57	8E-04	10	20	0	KLVVGGGGTGT	
48	48	2		Abbya gonypil	1014.571	10.4	41	2.9E-3	10	20	0	KLVVGGGGTGT	
49	49	2		Abbya gonypil	1014.571	10.4	57	8E-04	10	20	0	KLVVGGGGTGT	
50	50	2		Abbya gonypil	1014.571	10.4	41	2.9E-3	10	20	0	KLVVGGGGTGT	
51	51	2		Abbya gonypil	1014.571	10.4	57	8E-04	10	20	0	KLVVGGGGTGT	
52	52	2		Abbya gonypil	1014.571	10.4	41	2.9E-3	10	20	0	KLVVGGGGTGT	
53	53	2		Abbya gonypil	1014.571	10.4	57	8E-04	10	20	0	KLVVGGGGTGT	
54	54	2		Abbya gonypil	1014.571	10.4	41	2.9E-3	10	20	0	KLVVGGGGTGT	

MASCOT : MCP Guidelines without tears



Peptide match information is given on the second tab of the Excel report. MCP Guideline 2 states that you should give additional peptide level information for single-peptide based identifications. However, we include all of the approved peptide matches in this section. Here we must provide the peptide sequence noting any deviation from the expected cleavage pattern. Any modifications (fixed modifications are given as part of the database search conditions). We'll look in more detail at variable modifications in a short while as part of guideline 5. The precursor mass, charge and mass error observed have to be provided, as does the peptide score with any associated statistical information – here we provide the Mascot Ions Score and peptide e-value.

MS/MS fragmentation of KNLQYYDISAK
 Match to query 399 from mascot search mss-15052007-00007 (IF5 31-32_5006)
 From database (spare)c:\drivetest\listener\location\Chalkley_MCP_2005_4_1189_1193\w\HF5 5uLUCSF.wiff

Monoisotopic mass of neutral peptide Mr(calc): 1341.89
 Ions Score: 62 Expect: 0.0001655
 Matches (Bold Red): 15/104 fragment ions using 22 most intense peaks

#	b	b ⁺⁺	b ⁻	b ⁺⁻	b ⁰	b ⁰⁺⁺	Seq.	y	y ⁺⁺	y ⁻	y ⁺⁻	y ⁰	y ⁰⁺⁺	#
1	129.1	65.05	112.08	56.54			K							11
2	243.16	122.08	226.12	113.56			N	1214.61	607.81	1197.58	599.29	1196.59	598.8	10
3	356.23	178.62	339.2	170.1			L	1100.56	550.78	1083.54	542.27	1082.56	541.78	9
4	484.29	242.65	467.26	234.13			Q	987.48	494.24	970.45	485.73	969.47	485.24	8
5	647.35	324.18	630.32	315.67			Y	659.42	430.21	842.39	421.7	841.41	421.21	7
6	810.41	405.71	793.39	397.2			Y	696.36	348.68	679.33	340.17	678.35	339.68	6
7	925.44	463.22	908.41	454.71	907.43	454.22	D	533.29	267.15	518.27	258.64	515.28	258.14	5
8	1038.53	519.77	1021.5	511.25	1020.51	510.76	I	418.27	209.64	401.24	201.12	400.26	200.63	4
9	1126.66	563.28	1108.63	554.77	1107.55	554.28	S	305.18	153.09	288.16	144.58	287.17	144.09	3
10	1196.59	598.8	1179.57	590.29	1178.58	589.8	A	218.16	109.58	201.12	101.07			2
11							K	147.11	74.06	130.09	65.55			1

MASCOT : MCP Guidelines without tears © 2007 Matrix Science **MATRIX SCIENCE**

Guideline 2 also states that for any proteins identified by a single peptide, the MS/MS spectrum annotated with masses observed as well as fragment assignments should be included. These are supplied as the html files contained within the exported report zip file.

Molecular Cellular Proteomics publication guidelines

Guideline 3: Additional potentially valuable information

- Retention time of each peptide
- Observation of multiple charge states
- Multiple observations of the same peptide
- Flanking residues
- Start and end positions of peptides in proteins

MASCOT : MCP Guidelines without tears

© 2007 Matrix Science



The third guideline is concerned with any potentially valuable information. Any additional information that could be valuable can be supplied. This can include information such as the retention time of the peptides, charge state information, the residues flanking the peptide matches, or the position of the peptide matches in the protein.

Guideline 3: Additional information

Query ID	Query title	Rank	Observed(m/z)	Observed charge	M[exp]	M[calc]	Delta	Ions score	E-value	Start residue	End residue	Missed cleavages	Peptide sequence	Variable modifications	
1	msa.15652007.00003.GSP1_ASNGO GTP binding nuclear protein GSP118a	1	508.29299	2	1014.671	1014.571	5E-04	87	8E-4E-6	10	20	0	KLVLVGGDGGTK.T		
3	88 spectrum#51	1	508.29299	2	1014.577	1014.571	0.006	41	2.9E-3	10	20	0	KLVLVGGDGGTK.T		
7	msa.15052007.00003.MA1_YEAST Importin subunit alpha (Karyopherin beta)	1	373.23999	2	756.4654	756.4745	-0.01	43	1.4E-3	348	354	0	RLLLSPPK.E		
8	15 spectrum#40	1	373.23999	2	756.4654	756.4745	-0.01	43	1.4E-3	348	354	0	RLLLSPPK.E		
9	78 spectrum#10	1	482.760	2	963.9114	963.9277	-0.02	64	13.3E-6	391	398	0	KLLEVAEYK.T		
11	msa.15052007.00003.E1A_YEAST Elongation factor 1A (EF-1A)	1	382.21701	2	782.4196	782.4276	-0.01	38	4.0E-3	165	170	0	RFGKVK.E		
12	Query ID	Query title	Rank	Observed(m/z)	Observed charge	M[exp]	M[calc]	Delta	Ions score	E-value	Start residue	End residue	Missed cleavages	Peptide sequence	Variable modifications
13	16 spectrum#24	1	382.21701	2	782.4196	782.4276	-0.01	38	4.0E-3	165	170	0	RFGKVK.E		
14	55 spectrum#39	1	457.781006	2	913.5475	913.5597	-0.01	51	240.9E-6	429	437	0	RQTVAVGVK.S		
15	92 spectrum#91	1	513.30999	2	1024.605	1024.603	0.002	36	3.4E-3	254	264	0	KKGGGTVPPGR.V		
17	msa.15052007.00003.SYK1_YEAST Lysyl-tRNA synthetase, cytoplasmic	1	466.778015	2	929.5415	929.5588	-0.02	49	477.0E-6	301	308	0	RJAPLFLK.Q		
18	Query ID	Query title	Rank	Observed(m/z)	Observed charge	M[exp]	M[calc]	Delta	Ions score	E-value	Start residue	End residue	Missed cleavages	Peptide sequence	Variable modifications
19	62 spectrum#20	1	466.778015	2	929.5415	929.5588	-0.02	49	477.0E-6	301	308	0	RJAPLFLK.Q		
20	64 spectrum#13	1	470.75399	2	939.4334	939.5096	-0.01	36	5.2E-3	489	496	0	RFEVYVATK.E		
21	91 spectrum#106	1	512.288011	2	1022.567	1022.565	-0.01	34	8.8E-3	387	375	0	KKTEGVYK.V		
23	msa.15052007.00003.R516L_YEAST 40S ribosomal protein S19 A (S16A)	1	478.290009	2	954.5655	954.575	-0.01	60	16.2E-6	111	119	0	KKGVESPK.G		
24	Query ID	Query title	Rank	Observed(m/z)	Observed charge	M[exp]	M[calc]	Delta	Ions score	E-value	Start residue	End residue	Missed cleavages	Peptide sequence	Variable modifications
25	74 spectrum#126	1	478.290009	2	954.5655	954.575	-0.01	60	16.2E-6	111	119	0	KKGVESPK.G		
27	msa.15052007.00003.XP01_YEAST Exportin 1 (Chromosome region maintenance protein 1)	1	465.744001	2	925.4135	925.4182	-0	45	1.3E-3	31	40	0	KKAGELTK.F		
29	Query ID	Query title	Rank	Observed(m/z)	Observed charge	M[exp]	M[calc]	Delta	Ions score	E-value	Start residue	End residue	Missed cleavages	Peptide sequence	Variable modifications
30	61 spectrum#7	1	465.744001	2	925.4135	925.4182	-0	45	1.3E-3	31	40	0	KKAGELTK.F		
31	71 spectrum#31	1	476.770996	2	951.5274	951.5277	-0	40	1.8E-3	769	776	0	KLVVETYSK.A		
32	msa.15052007.00004.M5A_YEAST Importin beta 4 subunit (Karyopherin beta 4 subunit)	1	359.217987	2	716.4214	716.4432	-0.02	49	651.2E-6	719	725	0	KSVLLASK.V		
33	Query ID	Query title	Rank	Observed(m/z)	Observed charge	M[exp]	M[calc]	Delta	Ions score	E-value	Start residue	End residue	Missed cleavages	Peptide sequence	Variable modifications
34	6 spectrum#34	1	357.230011	2	713.4455	713.4687	-0.02	45	179.4E-6	1007	1022	0	KVIELLK.V		
35	9 spectrum#71	1	359.217987	2	716.4214	716.4432	-0.02	49	651.2E-6	719	725	0	KSVLLASK.V		
36	116 spectrum#18	1	422.227997	2	842.4414	842.481	-0.02	53	119.4E-6	57	64	0	KGLAGWEAR.K		
37	319 spectrum#55	1	520.23999	2	1039.445	1039.464	1E-03	46	1.1E-3	1077	1084	0	KGFQTEERK.H		
38	381 spectrum#55	1	556.333008	2	1110.651	1110.665	-0.01	46	367.4E-6	297	307	0	KLGPEITVAALK.V		
39	508 spectrum#354	1	629.814026	2	1257.614	1257.624	-0.01	43	1.3E-3	656	666	0	KFTVITGGYK.E		
40	537 spectrum#971	1	642.820021	2	1283.929	1283.935	-0.03	52	139.0E-6	604	614	0	KYGFDFAPPK.T		
41	559 spectrum#62	1	661.348027	2	1320.666	1320.692	-0.03	57	64.9E-6	949	960	0	KALVELLSAADQK.A		
42	561 spectrum#81	1	661.348026	2	1320.677	1320.693	-0.02	59	4.1E-3	84	85	0	KTSLLGTFSEPK.E		
43	582 spectrum#420	1	677.81979	2	1353.617	1353.635	-0.02	58	25.0E-6	550	561	0	RANTRFSTMTAK.A		
44	614 spectrum#36	1	761.377991	2	1520.741	1520.747	-0.01	58	49.0E-6	694	706	0	KVLMEDVDESYGR.E		
45	718 spectrum#281	1	687.840223	2	1378.85	1378.874	-0.02	52	162.0E-6	1028	1042	0	KLVGDFNSPVTNETPR.I		
46	720 spectrum#281	1	880.950809	2	1759.887	1759.874	0.013	44	1.2E-3	1028	1042	0	KLVGDFNSPVTNETPR.I		
47	47														
48	msa.15052007.00004.GSP1_ASNGO GTP binding nuclear protein GSP118a	1	508.29299	2	1014.671	1014.571	5E-04	87	8E-4E-6	10	20	0	KLVLVGGDGGTK.T		
49	Query ID	Query title	Rank	Observed(m/z)	Observed charge	M[exp]	M[calc]	Delta	Ions score	E-value	Start residue	End residue	Missed cleavages	Peptide sequence	Variable modifications
50	293 spectrum#171	1	508.29299	2	1014.577	1014.571	0.01	84	157.2E-6	10	20	0	KLVLVGGDGGTK.T		
51	295 spectrum#18	1	608.288011	2	1213.579	1213.598	-0.02	52	102.0E-6	140	149	0	KLVLVGGDGGTK.T		
52	457 spectrum#34	1	607.786997	2	1213.579	1213.598	-0.02	52	102.0E-6	140	149	0	KLVLVGGDGGTK.T		
53	459 spectrum#32	1	607.802003	2	1213.589	1213.596	-0.01	53	107.0E-6	140	149	0	KLVLVGGDGGTK.T		

MASCOT : MCP Guidelines without tears

© 2007 Matrix Science



Here we show the retention time of each peptide, the flanking residues multiple observations of the same peptide, and the start and end positions of the peptide in the protein. Because we're supplying details for all the peptide hits, we will also get any observation of multiple charge states.

Guideline 3: Additional information

EXP470500389 S. cerevisiae ICAT wash through experiment

Sample preparation conditions

Proteolytic digestion conditions (Time) 100 ul of each sample were diluted using 50 ul of 25 mM Ammonium Bicarb and subjected to proteolytic digestion using 2.5 mg of Trypsin at 37 degrees celcius for 4 hours

Chemical denaturation step 1 10 ug of each sample were diluted using 100 ul of ICAT - light (Applied Biosystems, Foster City, CA) and incubated at 21 degrees celcius for 1 hours

Chemical denaturation step 2 10 ug of each sample were diluted using 50 ul of ICAT - heavy (Applied Biosystems, Foster City, CA) and incubated at 21 degrees celcius for 1 hours

Purification step 1 Samples were purified by Elution. Samples were diluted using 10 ul of Biotin elution buffer.

Purification step 2 Samples were purified by Elution. Samples were diluted using 10 ul of Zypsi elution buffer. The purified samples were then washed 3 times in 100 ul of ddH2O.

Mass spec analysis MS analyses were carried out using a API QStar Pulsar I.

Peak picking parameters

Peak picking program MDRO (Mascot Distiller engine) 2.1.0.0

Database search parameters

Database search conditions set 1

Search engine	Mascot	2.2.1
Database	SwissProt	51.6
Database Size	257964	
Taxonomy	Fungi	
Database Size after Taxonomy	16473	4751
Peptide Mass Accuracy	60 ppm	
MS/MS Mass Accuracy	0.3 Da	
Maximum missed cleavages	2	
Fixed modifications		
Variable modifications	Acetyl (Protein N-term)	
Enzyme	Trypsin	
Resolution	NA	
Calibration	NA	
Exclusion of contaminant ions	NA	

Protein hit assignment criteria Proteins must have at least one bold significant peptide match with a p value < 0.01. Only significant peptide matches were approved.

SearchID	Search title	Source file	Protein Accession No	Protein Description	Mascot Score	Number matched peaks	Sequence Coverage	Number unique p
43	msa-16502007-00003	<fileaname> (F1 25-26_5001)	Ugpanec_dswar0-GSP1_A-ADGDO	GTP-binding nuclear prote	67.80229996	2	6.14	2
44	msa-16502007-00003	<fileaname> (F1 25-26_5001)	Ugpanec_dswar0-RNA1_YEAST	Import subunit alpha (oa	64.12	2	2.77	1
45	msa-16502007-00003	<fileaname> (F1 25-26_5001)	Ugpanec_dswar0-RS19A_YEAST	40S ribosomal protein S19	60.79	1	6.25	1
46	msa-16502007-00003	<fileaname> (F1 25-26_5001)	Ugpanec_dswar0-EF1A_YEAST	Elongation factor 1-alpha (54.44029996	3	5.68	3
47	msa-16502007-00003	<fileaname> (F1 25-26_5001)	Ugpanec_dswar0-SYWC_YEAST	Lysyl-tRNA synthetase, cy	52.63726658	3	4.23	2
48	msa-16502007-00003	<fileaname> (F1 25-26_5001)	Ugpanec_dswar0-XP01_YEAST	Exportin-1 (Chromosome I	49.29029996	2	1.48	2
49	msa-16502007-00004	<fileaname> (F2 26_5-26002)	Ugpanec_dswar0-EKSD_YEAST	Endonase 2 (EC 4.2.1.11) (64.63	2	4.58	2
50	msa-16502007-00004	<fileaname> (F2 26_5-26002)	Ugpanec_dswar0-RIB4_YEAST	Importin beta-4 subunit (K	237.17299996	13	11.32	13
51	msa-16502007-00004	<fileaname> (F2 26_5-26002)	Ugpanec_dswar0-GSP1_A-ADGDO	GTP-binding nuclear prote	192.60009999	4	9.81	4
52	msa-16502007-00004	<fileaname> (F2 26_5-26002)	Ugpanec_dswar0-EF1A_YEAST	Elongation factor 1-alpha (134.70149998	7	10.48	7
53	msa-16502007-00004	<fileaname> (F2 26_5-26002)	Ugpanec_dswar0-XP01_YEAST	Exportin-1 (Chromosome I	115.13816664	7	6.18	7
54	msa-16502007-00004	<fileaname> (F2 26_5-26002)	Ugpanec_dswar0-SM1_YEAST	Importin beta (SM1) (haryc	115.02009996	6	6.51	6

MASCOT : MCP Guidelines without tears

© 2007 Matrix Science



Because we tracked the experimental procedures through our Mascot Integra experiment, the system can automatically convert this information into an English description of some of these steps. For example, here we have a description of the proteolytic digestion conditions.

Guideline 4: Quantitative proteomics results

Query no	Protein	Ratio	Peptide rank	N	SD(geo)
54	VIT3_DROME	115/114	1	1.73	8.50
96	VIT3_DROME	115/114	1	1.89	1.19
114	VIT3_DROME	115/114	1	1.70	0.83
119	VIT3_DROME	115/114	1	1.69	1.07
134	VIT3_DROME	115/114	1	1.61	3.22
143	VIT3_DROME	115/114	1	1.04	0.47
148	VIT3_DROME	115/114	1	3.51	2.20
166	VIT3_DROME	115/114	1	1.56	4.06
172	VIT3_DROME	115/114	1	0.99	0.53
173	VIT3_DROME	115/114	1	0.88	0.61
229	VIT3_DROME	115/114	1	1.11	0.01

MASCOT : MCP Guidelines without tears

© 2007 Matrix Science



Guideline 4 covers Quantitative proteomics results. The experiment we just carried out did not contain any quantitation results so these results are from a different dataset. When you approve a protein hit with quantitation in Mascot Integra using one of the Mascot 2.2 methods which does not rely on the forthcoming Mascot Distiller Quantitation toolbox, the quantitation values and parameters used are also approved and stored in the database. Mascot Integra can report these results and settings in the MCP publication report.

Molecular Cellular Proteomics publication guidelines

Guideline 5: Studies focusing on posttranslational modifications require specialized methodology and documentation to assign the presence and the site(s) of modification.

- The sequence of the peptide used to make each such assignment
- The precursor mass and charge (not just m/z) observed
- The search engine score for this peptide
- An annotated and mass labelled spectra

MASCOT : MCP Guidelines without tears

© 2007 Matrix Science

MATRIX
SCIENCE

Guideline 5 is specific to modified peptides and requires that we supply additional information about post translationally modified peptides.

Guideline 5: Post-translational modifications

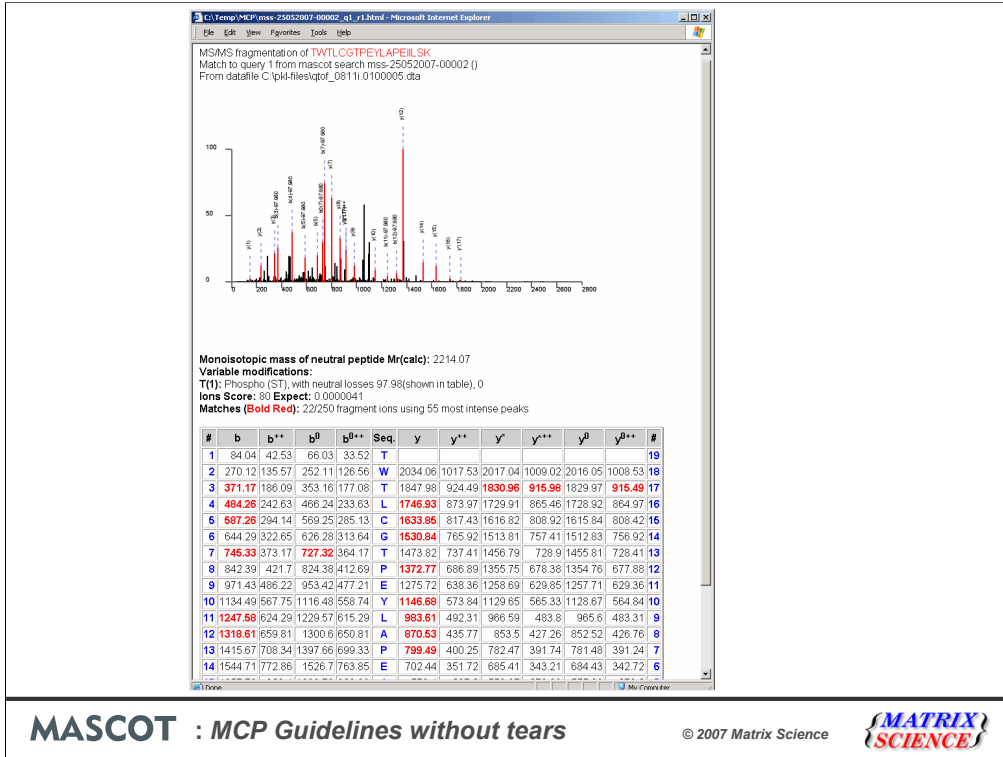
Rank	Observed charge	Molecular weight (kDa)	pI	Delta	Ion score	E-value	Start residue	End residue	Missed cleavages	Peptide sequence	Variable modifications
19	2	151.539	10.1	0.01	37	2.20	365	374	0	K.EIQDFQK.S	
20	2	1249.789	1249.764	0.005	35	0.35	144	151	0	R.LQAVYQK.Y	
21	2	1249.789	1249.764	0.005	35	0.35	144	151	0	R.LQAVYQK.Y	
22	2	1249.789	1249.764	0.005	35	0.35	144	151	0	R.LQAVYQK.Y	
23	2	1249.789	1249.764	0.005	35	0.35	144	151	0	R.LQAVYQK.Y	
24	2	1249.789	1249.764	0.005	35	0.35	144	151	0	R.LQAVYQK.Y	
25	2	1249.789	1249.764	0.005	35	0.35	144	151	0	R.LQAVYQK.Y	
26	2	1249.789	1249.764	0.005	35	0.35	144	151	0	R.LQAVYQK.Y	
27	2	1249.789	1249.764	0.005	35	0.35	144	151	0	R.LQAVYQK.Y	
28	2	1249.789	1249.764	0.005	35	0.35	144	151	0	R.LQAVYQK.Y	
29	2	1249.789	1249.764	0.005	35	0.35	144	151	0	R.LQAVYQK.Y	
30	2	1249.789	1249.764	0.005	35	0.35	144	151	0	R.LQAVYQK.Y	
31	2	1249.789	1249.764	0.005	35	0.35	144	151	0	R.LQAVYQK.Y	
32	2	1249.789	1249.764	0.005	35	0.35	144	151	0	R.LQAVYQK.Y	
33	2	1249.789	1249.764	0.005	35	0.35	144	151	0	R.LQAVYQK.Y	
34	2	1249.789	1249.764	0.005	35	0.35	144	151	0	R.LQAVYQK.Y	
35	2	1249.789	1249.764	0.005	35	0.35	144	151	0	R.LQAVYQK.Y	
36	2	1249.789	1249.764	0.005	35	0.35	144	151	0	R.LQAVYQK.Y	
37	2	1249.789	1249.764	0.005	35	0.35	144	151	0	R.LQAVYQK.Y	
38	2	1249.789	1249.764	0.005	35	0.35	144	151	0	R.LQAVYQK.Y	
39	2	1249.789	1249.764	0.005	35	0.35	144	151	0	R.LQAVYQK.Y	
40	2	1249.789	1249.764	0.005	35	0.35	144	151	0	R.LQAVYQK.Y	
41	2	1249.789	1249.764	0.005	35	0.35	144	151	0	R.LQAVYQK.Y	
42	2	1249.789	1249.764	0.005	35	0.35	144	151	0	R.LQAVYQK.Y	
43	2	1249.789	1249.764	0.005	35	0.35	144	151	0	R.LQAVYQK.Y	
44	2	1249.789	1249.764	0.005	35	0.35	144	151	0	R.LQAVYQK.Y	
45	2	1249.789	1249.764	0.005	35	0.35	144	151	0	R.LQAVYQK.Y	
46	2	1249.789	1249.764	0.005	35	0.35	144	151	0	R.LQAVYQK.Y	
47	2	1249.789	1249.764	0.005	35	0.35	144	151	0	R.LQAVYQK.Y	
48	2	1249.789	1249.764	0.005	35	0.35	144	151	0	R.LQAVYQK.Y	
49	2	1249.789	1249.764	0.005	35	0.35	144	151	0	R.LQAVYQK.Y	
50	2	1249.789	1249.764	0.005	35	0.35	144	151	0	R.LQAVYQK.Y	
51	2	2213.730	2214.000	-0.26	80	4.10E-6	196	214	0	R(t)W(t)LCGTPEYLAPEILSK.G	

MASCOT : MCP Guidelines without tears

© 2007 Matrix Science



When reporting these variable modifications, then the main additional requirement to the standard information reported for peptide matches, is that the site of modification be shown. This is shown in the report by changing the letter for the amino-acid residue to lower case – as for the oxidation of methionine predicted for this peptide. Where there is ambiguity as to the site of the modification, this must also be explicitly shown. Ambiguity is assumed by the system when there are multiple matches from Mascot to the same peptide and modification from the same query but with different modification locations where the ion score is above the significance threshold for the different modification locations. Ambiguity is shown in the report by putting brackets around the possible sites of the modification(s), as can be seen for the predicted Phosphorylation of one of these two serine residues on this peptide – the score for these other potential locations was below the significance threshold and so they are not reported.



As for single peptide identifications, the MCP guidelines require annotated spectra for the modified peptides are submitted. These are included in the exported zip file in the html files.

Molecular Cellular Proteomics publication guidelines

Guideline 7: Identical peptide sequences can be included in multiple unique protein sequences due to biological variation such as single amino acid variants, alternative splice forms, homologs, orthologs and paralogs:

- When assembling peptides into proteins and protein groups, authors should adhere to principles of parsimony, i.e., describe the minimum set of protein sequences that adequately accounts for all observed peptides.

Guideline 8: Include MS/MS data as supplemental material

- It is strongly encouraged (but not yet required) that all MS/MS spectra mentioned in the paper be submitted as supplemental material

MASCOT : MCP Guidelines without tears

© 2007 Matrix Science



We will skip guideline 6 for a few slides as it is specific for PMF's and go on to guidelines 7 and 8.

As mentioned earlier Peak lists produced during the automated data analysis with Mascot Daemon and saved as mzdata.xml files and are suitable for submitting as supplemental material if you choose to follow guideline 8.

Guideline 7 is concerned with which proteins are reported.

Molecular Cellular Proteomics publication guidelines

Guideline 6 Peptide Mass Fingerprinting

- Number of matched peaks
- Number of unmatched peaks
- Sequence coverage
- In addition to the score for the top match we must also show the score for the highest ranked hit to a non-homologous protein
- Spectra with matches marked
- Peaklists

MASCOT : MCP Guidelines without tears

© 2007 Matrix Science

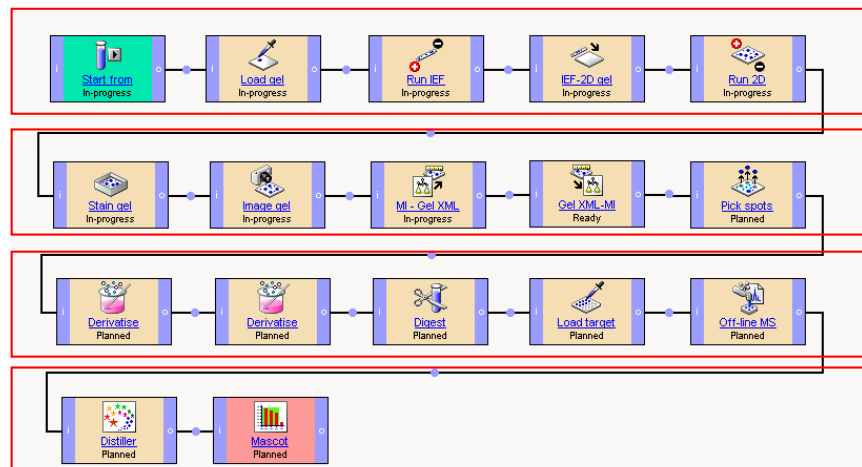


Coming back to Guideline 6, which is for peptide mass fingerprinting, the guideline requires the user to produce report that adheres to the guidelines that have already been detailed. Additionally the following information has to be provided: the number of matched and unmatched peaks; sequence coverage; and the score of the nearest non-homologous protein hit.

Depending on the redundancy of the database this may not necessarily be the second ranked protein hit.

We determine the score for the highest ranked hit to a non-homologous protein with BLAST cluster analysis.

Guideline 6: PMF reports



MASCOT : MCP Guidelines without tears

© 2007 Matrix Science

MATRIX
SCIENCE

Information is captured in the same way as a MS/MS based experiment. Shown here is the experimental workflow for a DIGE 2D gel analysis of *S pombe* cell lysate.

The work flow can be split into four sections:

Preparation and running of the 2D gel;

Staining, imaging, analysis and spot picking of the gel. Integra integrates with a number of different 2D gel analysis programs via xml export and import rather than providing gel analysis software;

Reduction, alkylation and digestion of the proteins and loading and MS analysis of the extracted tryptic peptides;

MS data analysis and database searching.

Once the workflow has been completed a MCP report can be exported.

Guideline 6: PMF reports

72
73 **Protein hit assignment criteria** PMF protein identifications were accepted if the expectation value (e-value) calculated by Mascot for the protein hit was below the 0.05 threshold.
74 The next best non-homologous protein hit was determined by using NCBI BLASTCluster with the following conditions: 40% identical residues, 50% minimum length coverage on one of the protein sequences
75

SearchID	Search title	Source file	Protein Accession No.	Protein Description	Mascot Score	Mascot e-value	Next best Mascot hit	Next best hit Mascot Score	Next best hit Mascot e-value
mss-02052007-00001	<filename>	(\spar\vspar\c_d\m\Mixture 1		Mixture from proteins: "	242	10.4E-2	SPB1_ENCUC	67.5	2.9E-3
77		UBA1_SCHPO	O94609	UBA1_SCHPO Ubiquitin-activating enzyme E1 1 (P9			(A)+ RNA transport protein 3) - Schizosaccharomyces pombe (Fission yeast)		
78		EF2_SCHPO	O14460	EF2_SCHPO Elongation factor 2 (EF-2) - Schizosaccharomyces pombe (Fission yeast)					
80		(\spar\vspar\c_d\m\CDC48_SCHPO		Cell division cycle prot	284	655.8E-27	HTF1_YEAST	35	4.14
81		(\spar\vspar\c_d\m\CDC48_SCHPO		Cell division cycle prot	284	655.8E-27	CYC_ROSNE	50.4	0.15
82		(\spar\vspar\c_d\m\CDC48_SCHPO		Cell division cycle prot	284	655.8E-27	PTB94_YEAST	48.1	0.26
83		(\spar\c_d\m\YHOF_ECOLI		Protein yhfF - Escheri	68.2	0.04	RNC_PSEF5	59.9	0.26
84		(\spar\c_d\m\YAG7_SCHPO		Hypothetical protein C1	53.3	0.04	YDH4_SCHPO	35	0.98

Number matched peaks	Number unmatched peaks	Sequence Coverage	Number unique peptides	Protein Mass	pl
72	67	0.00	40	0	0
		40.61		113960.23	5.11
		35.75		93796.81	5.02
36	32	47.85	34	90363.78	4.87
48	64	58.65	43	90363.78	4.87
50	66	58.75	41	90363.78	4.87
21	78	26.38	20	85352.06	5.92
11	88	32.52	11	45753.73	5.13

Export table of m/z values
* Required field

MASCOT : MCP Guidelines without tears © 2007 Matrix Science

As for the MS/MS report the experiment and database searches are selected for exporting. One big difference is that unapproved search results with a protein hit e-value filter are used rather than approved search results. As for the MS/MS report, additional information has to be entered by the user before the report can be generated. Once the report is complete Integra returns a link to a zip file containing the excel sheet and labeled PMF spectra.

The report is split over several sheets in an Excel workbook. The PMF report is very similar to the MS/MS report so I will only highlight differences required by Guideline 6.

On the first sheet below the search conditions there is a table of the protein hits. As you can see the report displays mixtures of proteins that can be detected in a spot. Significant expect scores are shown in red. Guideline 6 requires that the nearest non-homologous hit be reported. This is determined by clustering the protein hits with Blast.

Finally the number of matched and unmatched peaks along with sequence coverage and number of unique peptides are reported.

MCP Reports - status

- **PMF report completed (released as part of Mascot Integra 1.3)**
- **MS/MS still has work to be done**
 - Mechanism for exporting results without requiring protein approval
 - Quantitation
- **MS/MS report will be released as a patch soon.**

MASCOT : MCP Guidelines without tears

© 2007 Matrix Science



The MCP compliant report for PMF experiments has been completed and is part of Mascot Integra 1.3. The MS/MS report still has some work to be completed. 1) A mechanism for exporting MS/MS experiments where the user does not wish to carryout manual approval – this has implications for MCP guideline 7. 2) A mechanism to allow the user to capture additional information required for MCP guideline 4 (Quantitative Proteomics), and to support quantitation methods that require the use of the forthcoming Mascot Distiller Quantitation toolbox. This work has almost been completed and we will be releasing the MCP MS/MS compliant report in the near future as an update to Mascot Integra 1.3

Reporting standards

MIAPE

Taylor CF. Minimum Reporting Requirements for Proteomics: A MIAPE Primer. *Proteomics*. 2006 Sep;6 Suppl 2:39-44.

Community Consultation on standards papers for publication in Nature Biotechnology:

<http://www.nature.com/nbt/consult/index.html>

PRIDE (PRoteomics IDentifications) database

Martens L, *et al* PRIDE: the proteomics identifications database. *Proteomics*. 2005 Aug;5(13):3537-45

MASCOT : MCP Guidelines without tears

© 2007 Matrix Science



The MCP guidelines specify the types of information that should be reported but not the format for the information.

There are a number of reporting standards in development and in a few cases such as mzData.xml already in use.

The Minimum Information About a Proteomics Experiment or MIAPE is a reporting standard backed by the Proteomics Standards Initiative of HUPO. It specifies a format and controlled vocabulary for reporting Proteomics experiments. MIAPE and a number of other standards formats have recently been under public review prior to publication in Nature Biotechnology. I am afraid it is too late to make any input now but you can still look at the MIAPE documents on the Nature website.

There are also a number of public data repositories which can store raw data, peaklists and experimental information, one of which is PRIDE. PRIDE is hosted by the European Bioinformatics Institute and is a standards compliant, public data repository for proteomics data. Experimental data is submitted in a standardized XML format using a predefined ontology. It is one of a number of raw data repositories but as it also stores data about the experimental conditions along with the associated MS data is more valuable than a data only repository.

MAIPE, PSI standards and PRIDE

Reporting Standards (MAIPE)

- FUGE, GelML, GellInfoML, spML, mzData (v1.05), analysisXML, MIF
- The PSI mzData.xml standard for MS peaklists is already well supported.

The PRIDE database was designed to provide a common data exchange format and repository to support proteomics literature publications.

- An export to PRIDE.xml feature is currently in development.

MASCOT : MCP Guidelines without tears

© 2007 Matrix Science



We already support mzXML and mzData and are actively involved in the development of MAIPE standards.

The PRIDE database was designed to provide a common data exchange format and repository to support proteomics literature publications. It can be used as a central location to provide anonymous access by reviewers to experiment results. On publication of the manuscript the author can choose to release the data to the public.

As I mentioned in the previous slide PRIDE use there own standardized XML format for importing the experiment results, but will also support analysisXML in the future.

We are currently working on a direct to PRIDE export feature.

Acknowledgments

Robert Chalkley *et al* for making the data used in the LCMS/MS example report publicly available.

Mol Cell Proteomics. 2005 Aug;4(8):1189-93

Dr Kathryn Lilley for permission to use the iTraQ dataset, which was acquired as part of a *Drosophila* study.

Mark Weeks *et al* for the *S. pombe* 2D gel data set.

Proteomics. 2005 Apr;5(6):1669-85

The reports we have shown used data from the following studies and we are grateful to the authors for making the data either publicly available or giving us permission to use it.