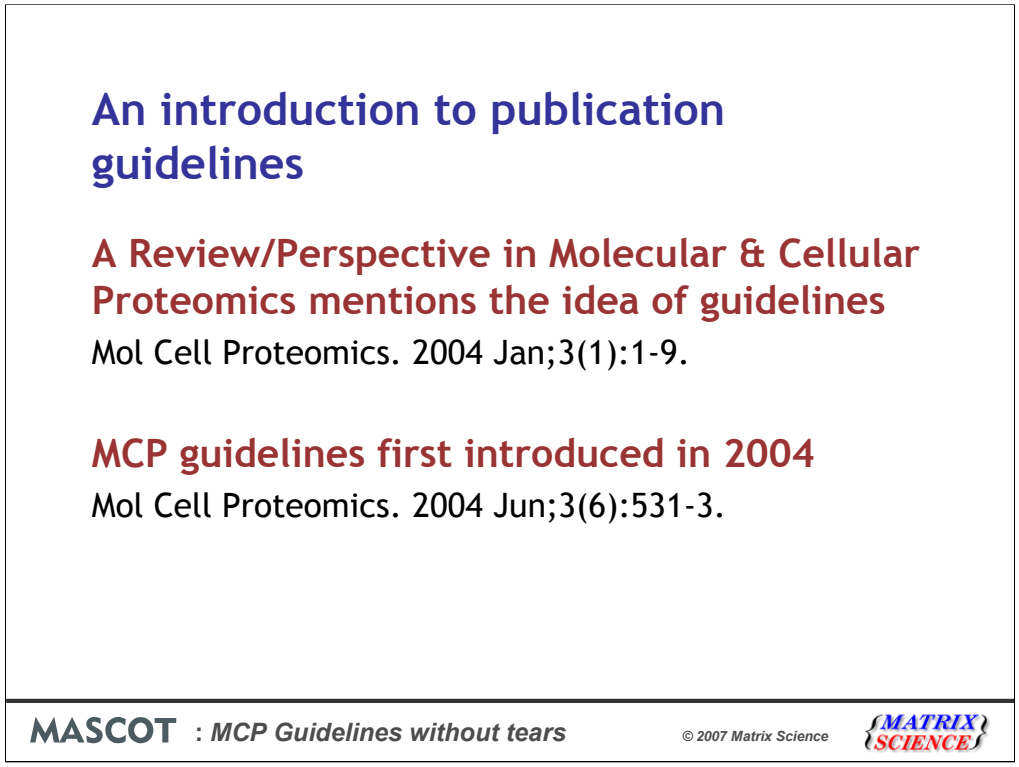
MASCOT*Integra*

MCP Guidelines without tears



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.

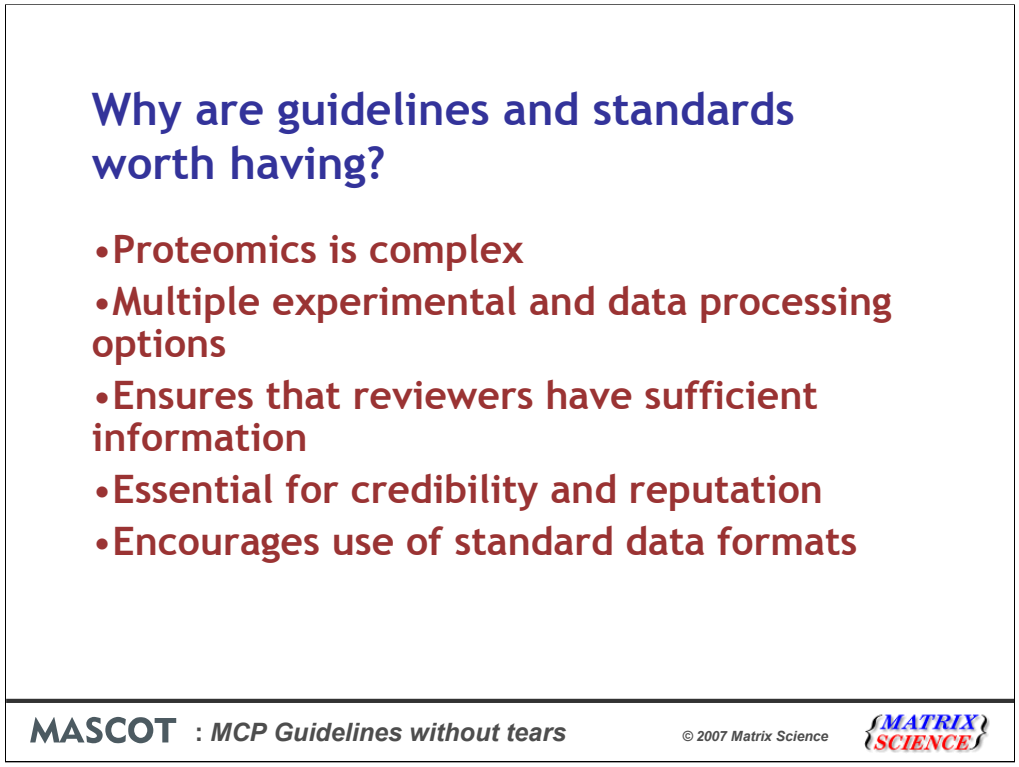
http://www.mcponline.org/misc/ParisReport_Final.shtml

MCP are not the only journal to provide publication guidelines

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

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.

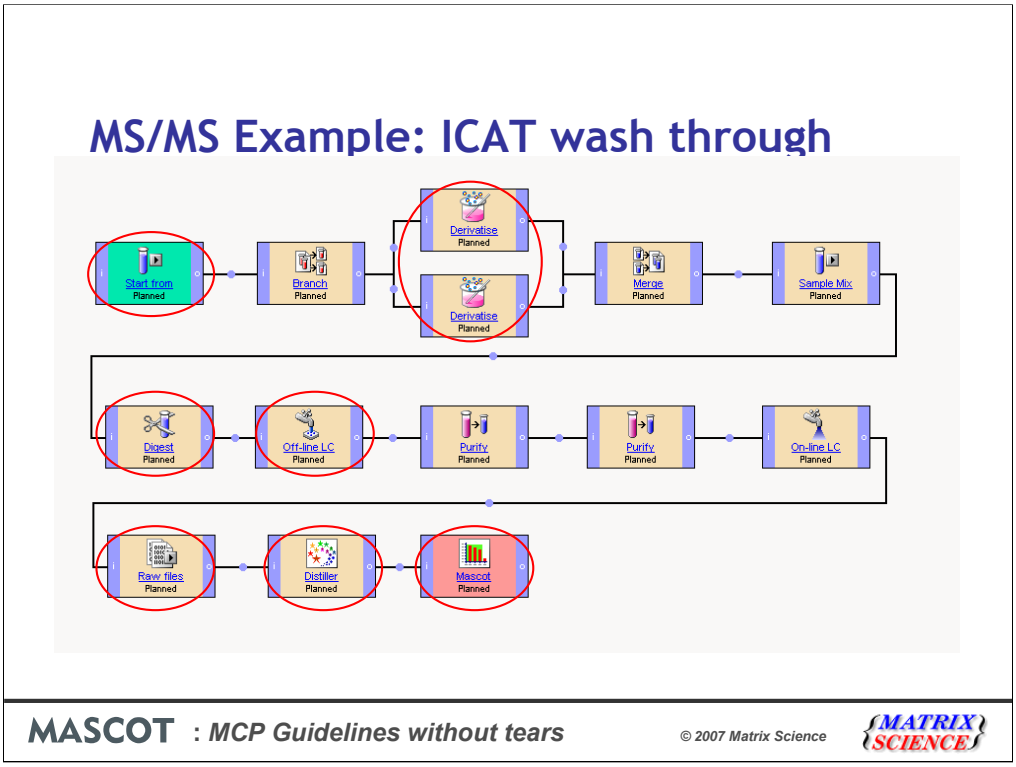


Proteomics is complex, with many options available during both the experimental stages and data analysis. The parameters used need to be clearly documented. For a reviewer or reader to fully comprehend the results certain information needs to be presented. Guidelines help by defining what should be included.

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.

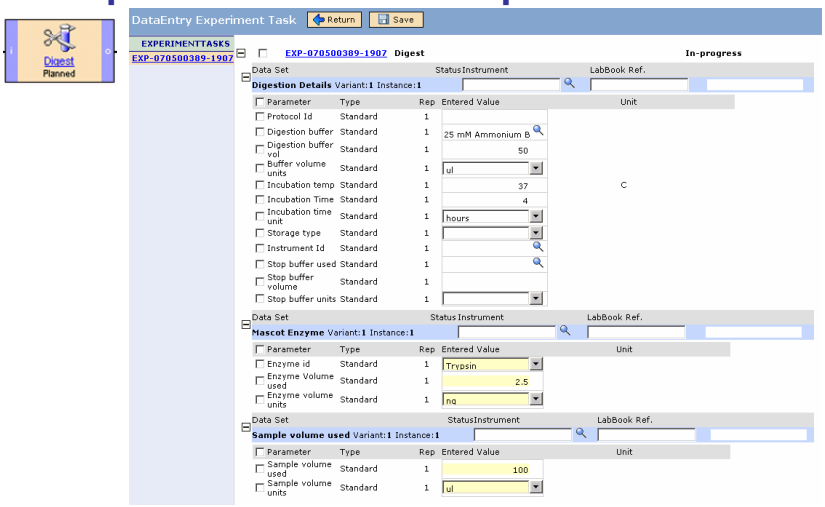
Reporting standards can be more stringent than guidelines in that they normally define a vocabulary and format for the information. However a standard might not require all the information that guidelines recommend.

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

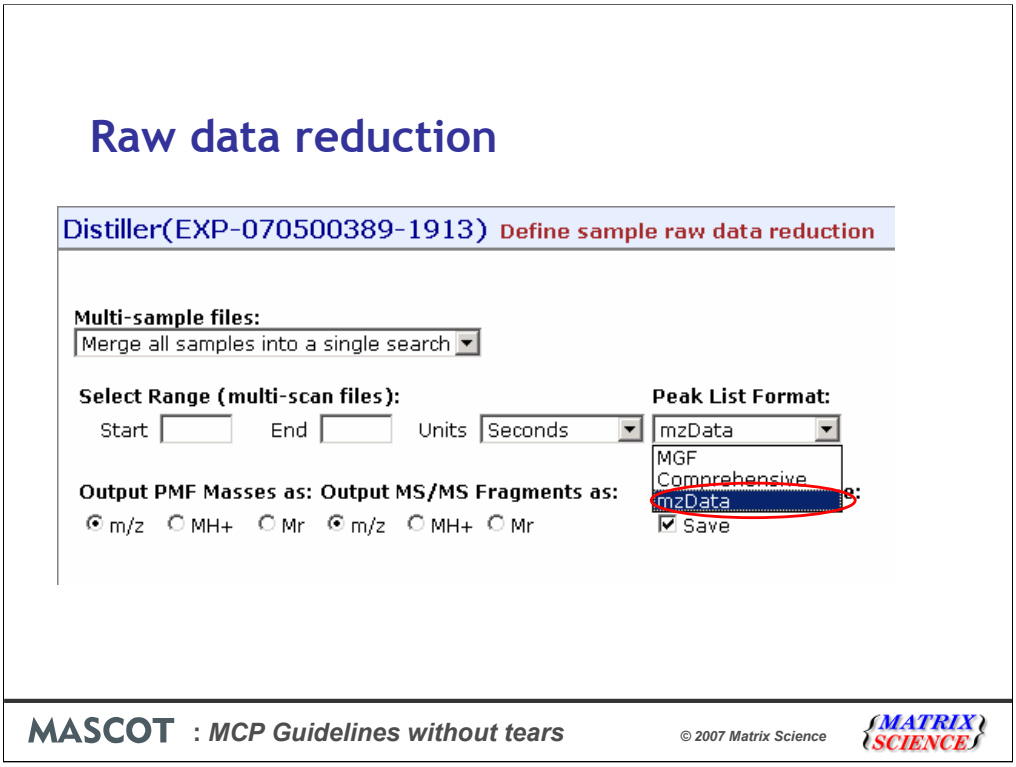


Here we have modelled an MS/MS experiment in Mascot Integra which supports the ICAT procedure, but where we are going to analyse the peptides that did not bind to the avidin column. Protein extracts from G1 and M-phase arrested *S.cerevisiae* cells were labelled with the ICAT light and heavy reagents 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



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.

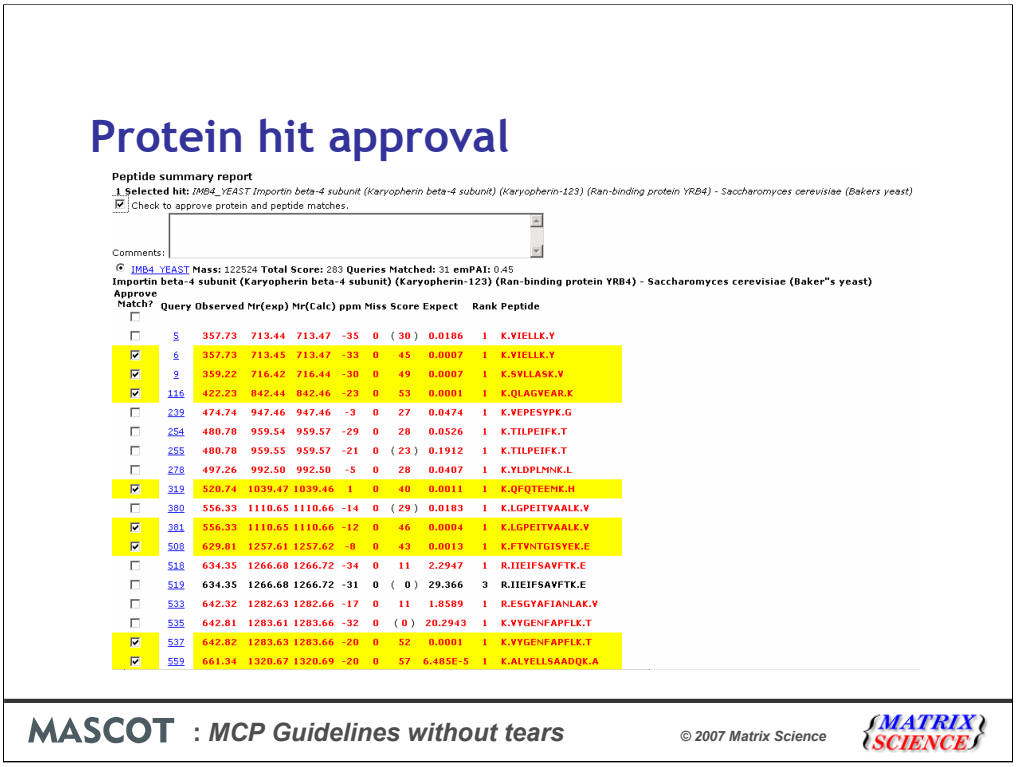


During the raw data reduction step using Mascot Distiller we're going to export the peaklists as mzData files. mzData is an information rich peaklist file format specified by the Proteomics Standards Initiative workgroup (PSI) of HUPO. If you're going to use Mascot Integra (or any other system) to generate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

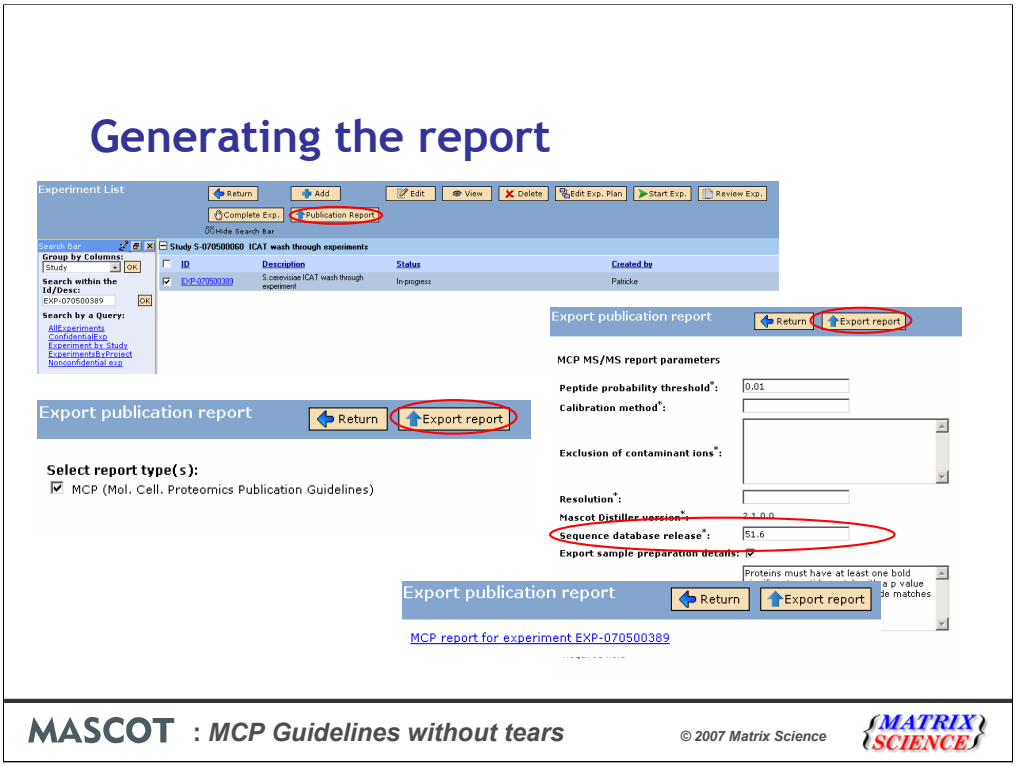
1.It is a requirement of the mzData format that the software (and version number) used to generate the peaklist is included in the file. Also some of the peak detection parameters are often included. We need this information to generate the MCP report and so using the mzData format removes the requirement 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.

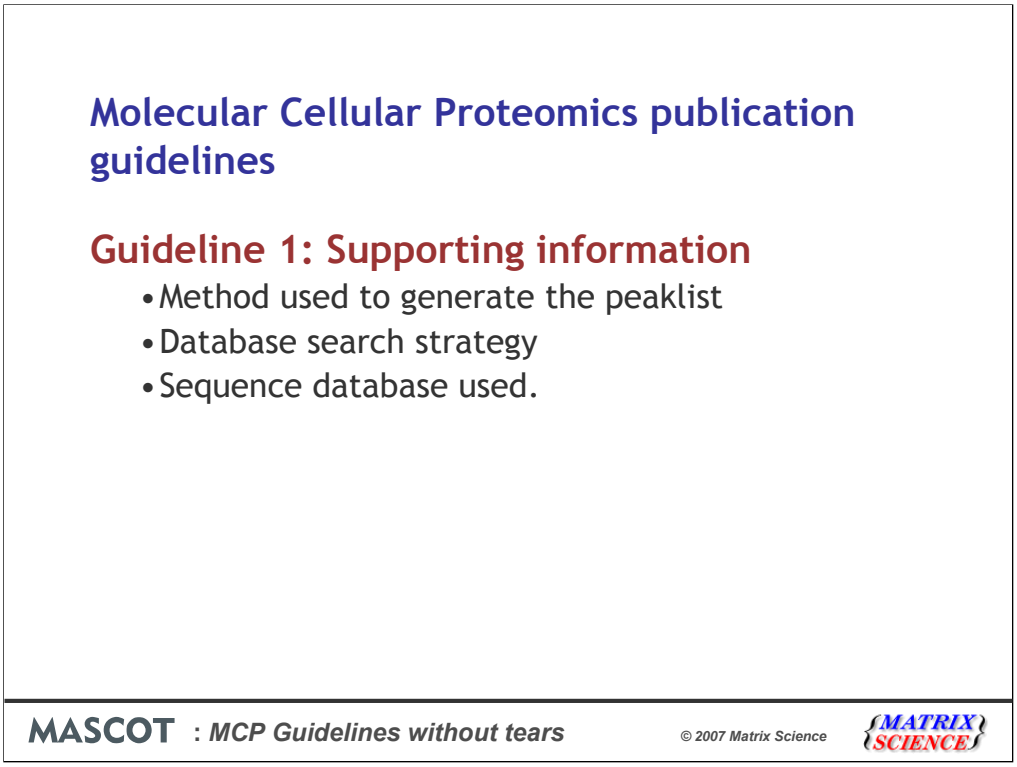
3.If you choose to submit your experiment to PRIDE, mzData is also the required format for submitting peaklists, so generating the peaklists as mzData files here could save you time later on.



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.



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.



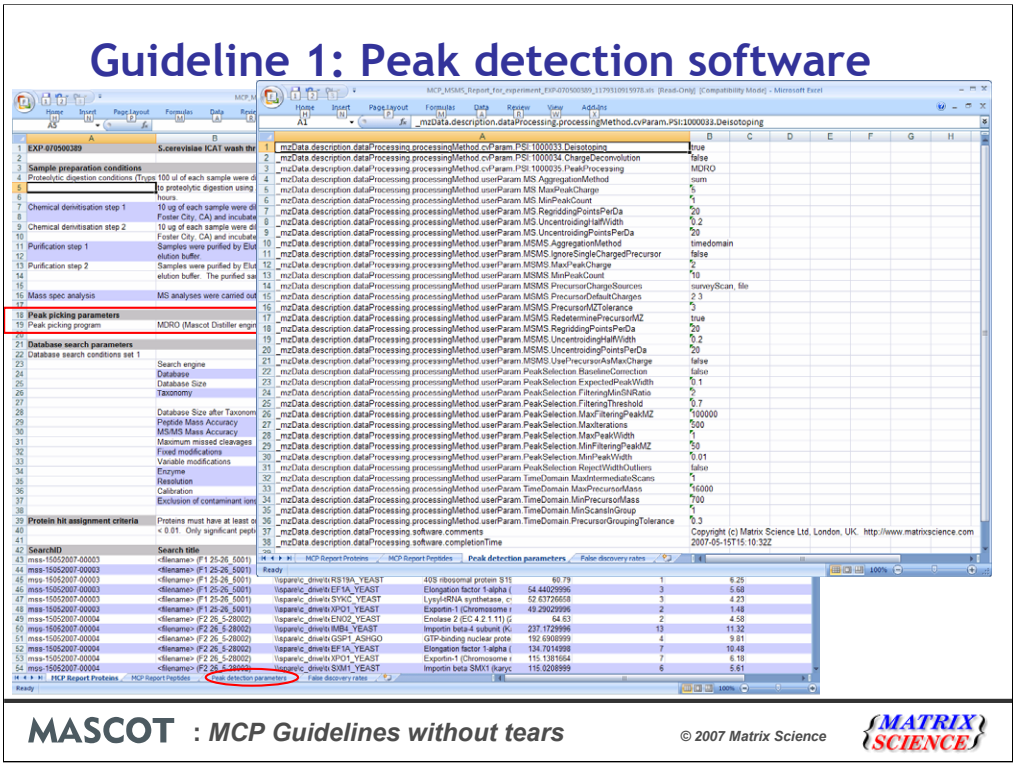
OK, now lets review the Molecular Cellular Proteomics publication guidelines to see what information we have to provide, and then look at the report from Mascot Integra to see how we've supported that requirement.

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

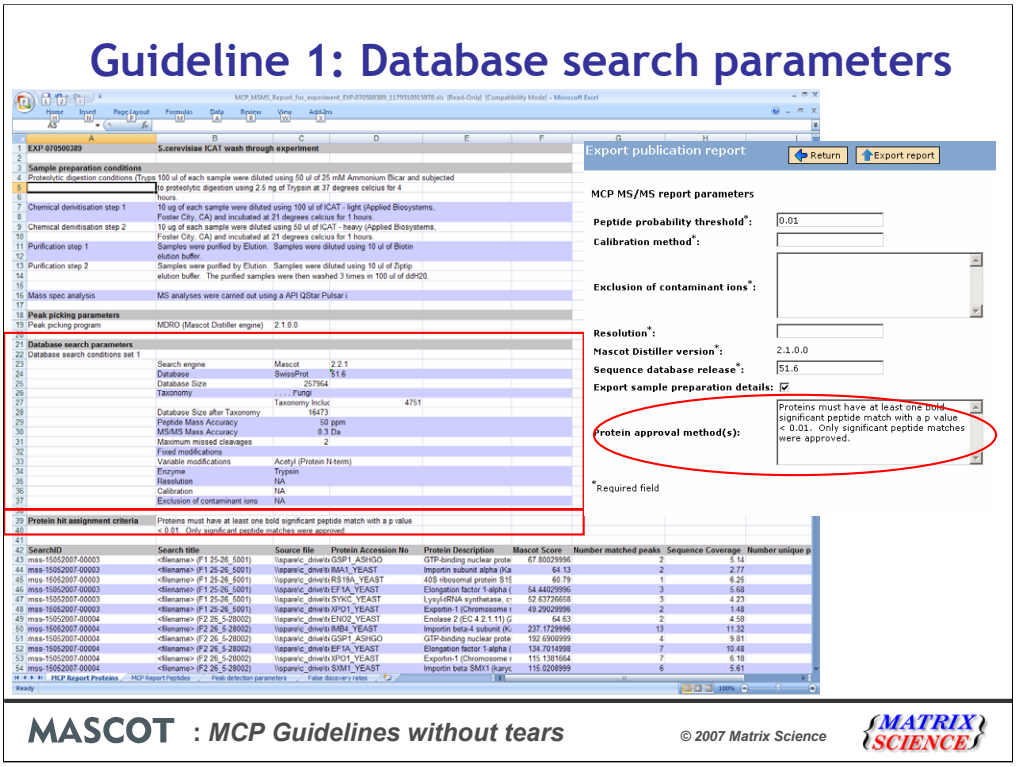
The method and/or program (including version number) used to create the "peak list" from the raw data and the parameters used in the creation of this peak list

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.

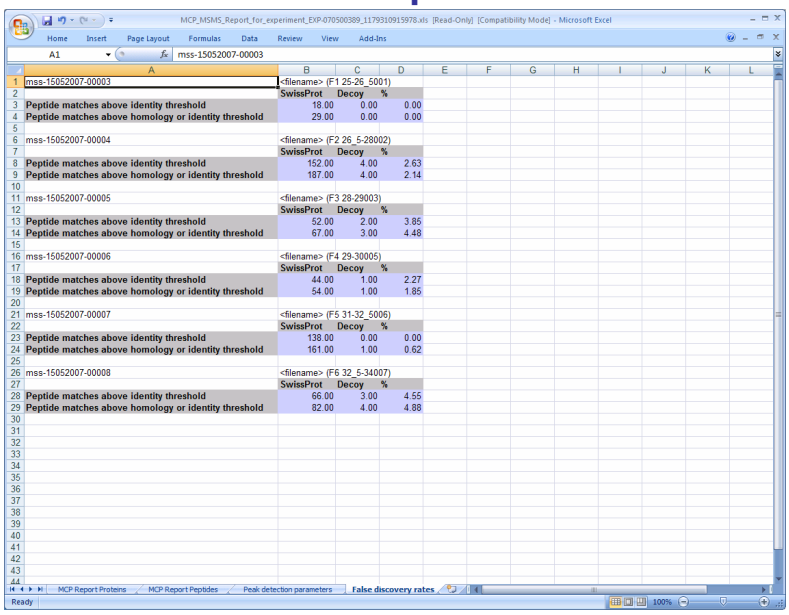


The program and version number used to create the peaklist from the raw data are shown here on the first tab in the Excel report. We must also supply the parameters used in the creation of the peak list. This information has been put in a separate tab - 'Peak detection parameters'. Here we can clearly 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 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



	A	B	C	D
1	mss-15052007-00003	<filename> (F1 25-26_5001)		
2		SwissProt	Decoy	%
3	Peptide matches above identity threshold	18.00	0.00	0.00
4	Peptide matches above homology or identity threshold	29.00	0.00	0.00
5				
6	mss-15052007-00004	<filename> (F2 26_5-28002)		
7		SwissProt	Decoy	%
8	Peptide matches above identity threshold	152.00	4.00	2.63
9	Peptide matches above homology or identity threshold	187.00	4.00	2.14
10				
11	mss-15052007-00005	<filename> (F3 28-29003)		
12		SwissProt	Decoy	%
13	Peptide matches above identity threshold	52.00	2.00	3.85
14	Peptide matches above homology or identity threshold	67.00	3.00	4.48
15				
16	mss-15052007-00006	<filename> (F4 29-30005)		
17		SwissProt	Decoy	%
18	Peptide matches above identity threshold	44.00	1.00	2.27
19	Peptide matches above homology or identity threshold	54.00	1.00	1.85
20				
21	mss-15052007-00007	<filename> (F5 31-32_5006)		
22		SwissProt	Decoy	%
23	Peptide matches above identity threshold	138.00	0.00	0.00
24	Peptide matches above homology or identity threshold	161.00	1.00	0.62
25				
26	mss-15052007-00008	<filename> (F6 32_5-34007)		
27		SwissProt	Decoy	%
28	Peptide matches above identity threshold	66.00	3.00	4.55
29	Peptide matches above homology or identity threshold	82.00	4.00	4.88

MASCOT : *MCP Guidelines without tears* *© 2007 Matrix Science* 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.

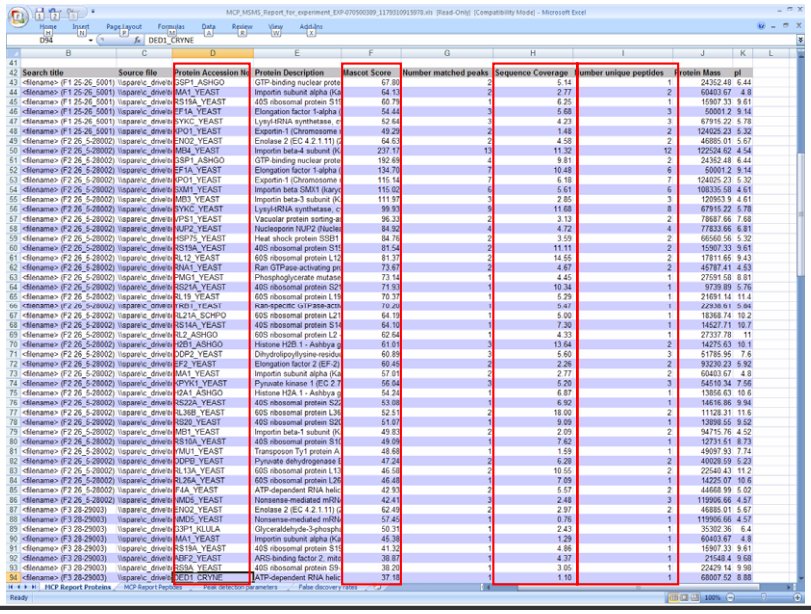
The second guideline specifies the protein information that should be reported. This includes details about the protein accession and database source

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

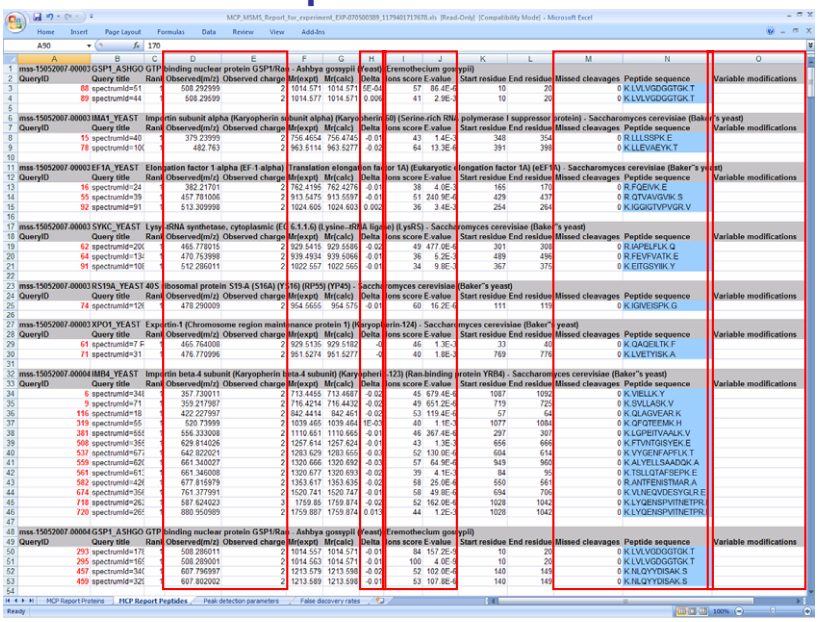


Search title	Source file	Protein Accession No	Protein Description	Mascot Score	Number matched peaks	Sequence Coverage	Number unique peptides	Protein Mass	pI
<filename> (F1 25-26_5001)	\\spare\c_drive\b	GSP1_ASHGO	GTP-binding nuclear prote	67.80	2	5.14	1	24352.48	6.44
<filename> (F1 25-26_5001)	\\spare\c_drive\b	MA1_YEAST	Importin subunit alpha (Ka	64.13	2	2.77	2	60403.67	4.8
<filename> (F1 25-26_5001)	\\spare\c_drive\b	RS19A_YEAST	40S ribosomal protein S19	60.79	1	6.25	1	15907.33	9.61
<filename> (F1 25-26_5001)	\\spare\c_drive\b	EF1A_YEAST	Elongation factor 1-alpha (	54.44	3	5.68	3	50001.2	9.14
<filename> (F1 25-26_5001)	\\spare\c_drive\b	SYKC_YEAST	Lysyl-tRNA synthetase, c	52.64	3	4.23	3	67915.22	5.78
<filename> (F1 25-26_5001)	\\spare\c_drive\b	XPO1_YEAST	Exportin-1 (Chromosome	49.29	2	1.48	2	124025.23	5.32
<filename> (F2 26_5-28002)	\\spare\c_drive\b	ENO2_YEAST	Enolase 2 (EC 4.2.1.11) (	64.63	2	4.58	2	46885.01	5.67
<filename> (F2 26_5-28002)	\\spare\c_drive\b	MB4_YEAST	Importin beta-4 subunit (K	237.17	13	11.32	12	122524.62	4.54
<filename> (F2 26_5-28002)	\\spare\c_drive\b	GSP1_ASHGO	GTP-binding nuclear prote	192.68	4	9.81	2	24362.48	6.44
<filename> (F2 26_5-28002)	\\spare\c_drive\b	EF1A_YEAST	Elongation factor 1-alpha (	134.70	7	10.48	6	50001.2	9.14
<filename> (F2 26_5-28002)	\\spare\c_drive\b	XPO1_YEAST	Exportin-1 (Chromosome	115.14	7	6.18	7	124025.23	5.32
<filename> (F2 26_5-28002)	\\spare\c_drive\b	IMA1_YEAST	Importin beta SMX1 (karyo	115.02	6	5.61	6	108335.58	4.61
<filename> (F2 26_5-28002)	\\spare\c_drive\b	MB3_YEAST	Importin beta-3 subunit (K	111.97	3	2.85	3	120953.9	4.61
<filename> (F2 26_5-28002)	\\spare\c_drive\b	SYKC_YEAST	Lysyl-tRNA synthetase, c	99.93	9	11.68	8	67915.22	5.78
<filename> (F2 26_5-28002)	\\spare\c_drive\b	VPS1_YEAST	Vacuolar protein sorting-a	96.33	2	3.13	2	78687.66	7.68
<filename> (F2 26_5-28002)	\\spare\c_drive\b	NUP2_YEAST	Nucleoporin NUP2 (Nucle	84.92	4	4.72	4	77833.66	6.81
<filename> (F2 26_5-28002)	\\spare\c_drive\b	HSP75_YEAST	Heat shock protein SSB1	84.76	2	3.59	2	66560.56	5.32
<filename> (F2 26_5-28002)	\\spare\c_drive\b	RS19A_YEAST	40S ribosomal protein S19	81.54	2	11.11	2	15907.33	9.61
<filename> (F2 26_5-28002)	\\spare\c_drive\b	RL12_YEAST	60S ribosomal protein L12	81.37	2	14.55	2	17811.66	9.43
<filename> (F2 26_5-28002)	\\spare\c_drive\b	RNA1_YEAST	Ran GTPase-activating pro	73.67	2	4.67	2	45787.41	4.53
<filename> (F2 26_5-28002)	\\spare\c_drive\b	PMG1_YEAST	Phosphoglycerate mutase	73.14	1	4.45	1	27591.58	8.81
<filename> (F2 26_5-28002)	\\spare\c_drive\b	RS25A_YEAST	40S ribosomal protein S25	71.93	1	10.34	1	9739.89	5.76
<filename> (F2 26_5-28002)	\\spare\c_drive\b	RL19_YEAST	60S ribosomal protein L19	70.37	1	5.29	1	21691.14	11.4
<filename> (F2 26_5-28002)	\\spare\c_drive\b	KRD1_YEAST	Ran-specific GTPase-acti	70.26	1	5.47	1	22936.61	5.84
<filename> (F2 26_5-28002)	\\spare\c_drive\b	RL21A_SCHPO	60S ribosomal protein L21	64.19	1	5.00	1	18368.74	10.2
<filename> (F2 26_5-28002)	\\spare\c_drive\b	RS14A_YEAST	40S ribosomal protein S14	64.10	1	7.30	1	14527.71	10.7
<filename> (F2 26_5-28002)	\\spare\c_drive\b	RL2_ASHGO	60S ribosomal protein L2	62.64	1	4.33	1	27337.78	11
<filename> (F2 26_5-28002)	\\spare\c_drive\b	H2B1_ASHGO	Histone H2B.1 - Ashbya g	61.81	3	13.64	2	14275.63	10.1
<filename> (F2 26_5-28002)	\\spare\c_drive\b	DDP2_YEAST	Dihydrolipoyllysine-resido	60.89	3	5.60	3	51786.95	7.6
<filename> (F2 26_5-28002)	\\spare\c_drive\b	EF2_YEAST	Elongation factor 2 (EF-2)	60.45	2	2.26	2	93230.23	5.92
<filename> (F2 26_5-28002)	\\spare\c_drive\b	IMA1_YEAST	Importin subunit alpha (Ka	57.81	2	2.77	2	60403.67	4.8
<filename> (F2 26_5-28002)	\\spare\c_drive\b	KPYK1_YEAST	Pyruvate kinase 1 (EC 2.7.	56.54	3	5.29	3	54519.34	7.56
<filename> (F2 26_5-28002)	\\spare\c_drive\b	H2A1_ASHGO	Histone H2A.1 - Ashbya g	54.24	1	6.87	1	13856.63	10.6
<filename> (F2 26_5-28002)	\\spare\c_drive\b	RS22A_YEAST	40S ribosomal protein S22	53.08	1	6.92	1	14616.86	9.94
<filename> (F2 26_5-28002)	\\spare\c_drive\b	RL36B_YEAST	60S ribosomal protein L36	52.51	2	18.00	2	11128.31	11.6
<filename> (F2 26_5-28002)	\\spare\c_drive\b	RS20_YEAST	40S ribosomal protein S20	51.67	1	9.09	1	12898.55	9.52
<filename> (F2 26_5-28002)	\\spare\c_drive\b	IMB1_YEAST	Importin beta-1 subunit (K	49.83	2	2.09	2	94715.76	4.52
<filename> (F2 26_5-28002)	\\spare\c_drive\b	RS10A_YEAST	40S ribosomal protein S10	49.09	1	7.62	1	12731.61	8.73
<filename> (F2 26_5-28002)	\\spare\c_drive\b	YMU1_YEAST	Transposon Ty1 protein A	48.68	1	1.59	1	49097.93	7.74
<filename> (F2 26_5-28002)	\\spare\c_drive\b	ODPB_YEAST	Pyruvate dehydrogenase E	47.24	2	6.28	2	40028.59	5.23
<filename> (F2 26_5-28002)	\\spare\c_drive\b	RL13A_YEAST	60S ribosomal protein L13	46.58	2	10.55	2	22540.47	11.2
<filename> (F2 26_5-28002)	\\spare\c_drive\b	RL26A_YEAST	60S ribosomal protein L26	46.48	1	7.09	1	14225.97	10.6
<filename> (F2 26_5-28002)	\\spare\c_drive\b	F4A_YEAST	ATP-dependent RNA helic	42.93	2	5.57	2	44668.99	5.02
<filename> (F2 26_5-28002)	\\spare\c_drive\b	NMD5_YEAST	Nonsense-mediated mRN	42.41	3	2.48	3	119906.66	4.57
<filename> (F3 28-29003)	\\spare\c_drive\b	ENO2_YEAST	Enolase 2 (EC 4.2.1.11) (	62.49	2	2.97	2	46885.01	5.67
<filename> (F3 28-29003)	\\spare\c_drive\b	NMD5_YEAST	Nonsense-mediated mRN	57.45	1	0.76	1	119906.66	4.57
<filename> (F3 28-29003)	\\spare\c_drive\b	G3P1_KLULA	Glyceraldehyde-3-phosph	50.31	1	2.43	1	35302.36	6.4
<filename> (F3 28-29003)	\\spare\c_drive\b	MA1_YEAST	Importin subunit alpha (Ka	46.38	1	1.29	1	60403.67	4.8
<filename> (F3 28-29003)	\\spare\c_drive\b	RS19A_YEAST	40S ribosomal protein S19	41.32	1	4.86	1	15907.33	9.61
<filename> (F3 28-29003)	\\spare\c_drive\b	ABF2_YEAST	ARS-binding factor 2, mito	38.87	1	4.37	1	21548.4	9.68
<filename> (F3 28-29003)	\\spare\c_drive\b	RS9A_YEAST	40S ribosomal protein S9	38.26	1	3.05	1	22429.14	9.98
<filename> (F3 28-29003)	\\spare\c_drive\b	DED1_CRYNE	ATP-dependent RNA helic	37.18	1	1.10	1	68007.52	8.88

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

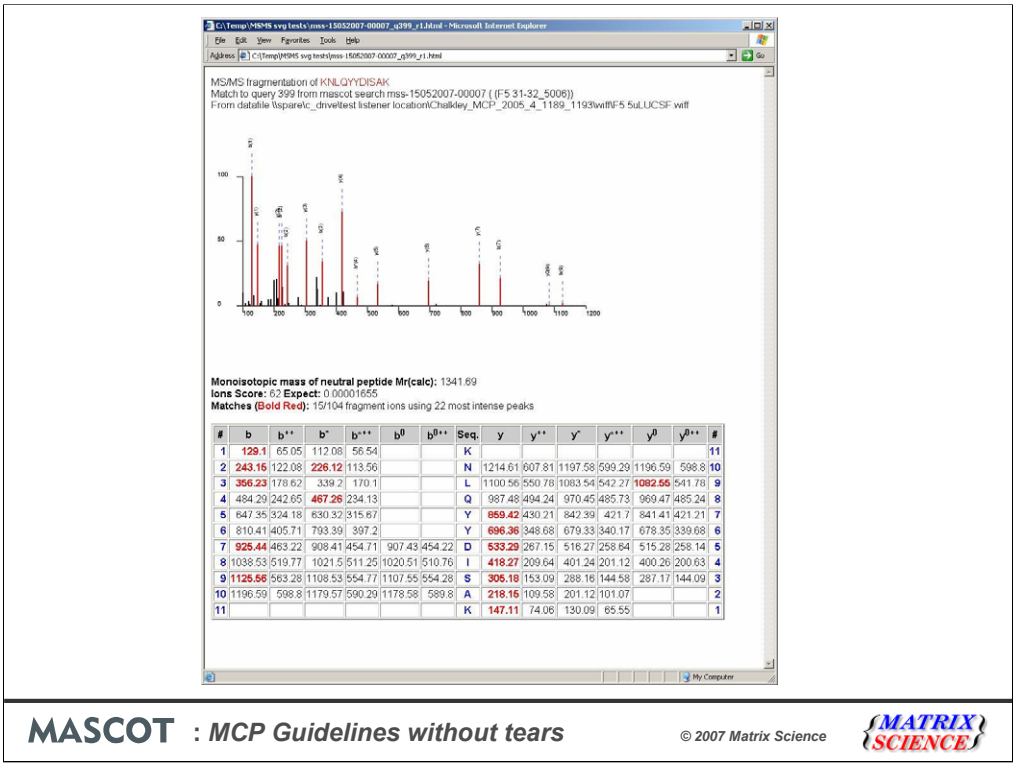
On the first tab of the exported report, below the search parameters, we provide the protein hit information. Here we have the accession number. The database source is provided as part of the search parameters. The protein scores are recalculated to include only the peptide matches that we approved. 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



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

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.



#	b	b++	b*	b*++	b0	b0++	Seq.	y	y++	y*	y*++	y0	y0++	#
1	129.1	65.05	112.08	56.54			K							11
2	243.15	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	1082.55	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	859.42	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	516.27	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	1125.56	563.28	1108.53	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	218.15	109.58	201.12	101.07			2
11							K	147.11	74.06	130.09	65.55			1

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.

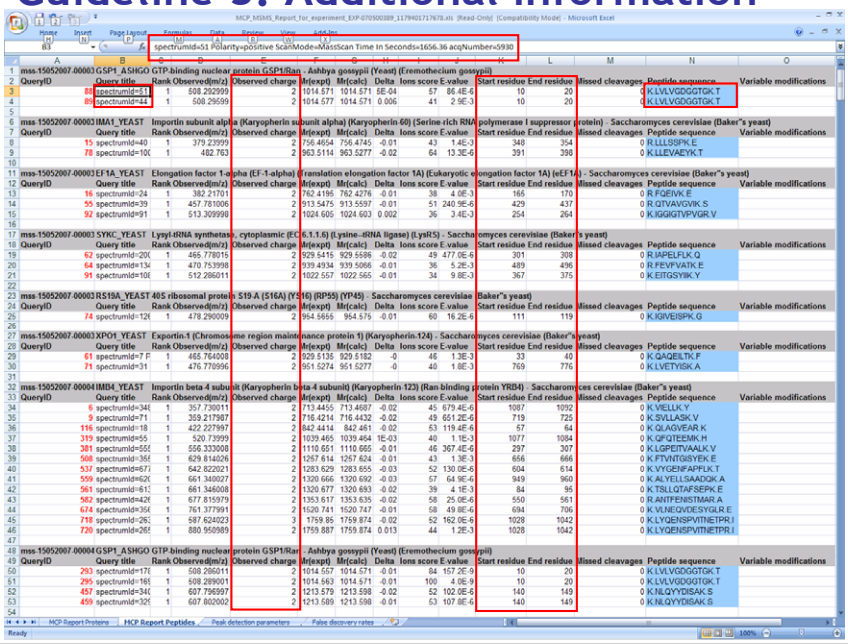
Molecular Cellular Proteomics publication guidelines

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

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.

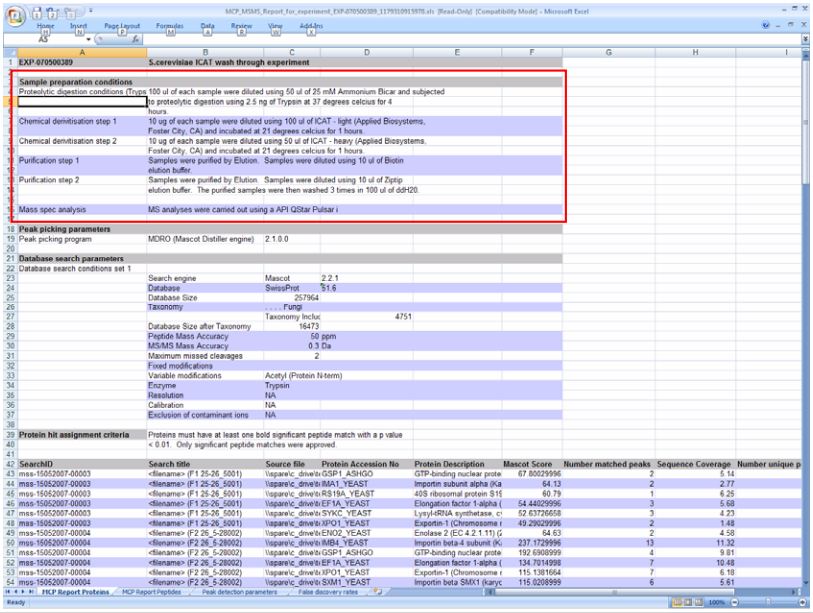
Guideline 3: Additional information



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



EXP.070500389	S.cerevisiae ICAT wash through experiment	
Sample preparation conditions		
Proteolytic digestion conditions (Trypsi	100 ul of each sample were diluted using 50 ul of 25 mM Ammonium Bicar and subjected to proteolytic digestion using 2.5 ng of Trypsin at 37 degrees celcius for 4 hours	
Chemical derivitisation 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 derivitisation 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 Ziptip elution buffer. The purified samples were then washed 3 times in 100 ul of ddH20	
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
	Taxonomy Inclus	4751
	Database Size after Taxonomy	16473
	Peptide Mass Accuracy	50 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
mss-15052007-00003	<filename> (F1 25-26_5001)	\\spare\c_drive\b\GSP1_ASHGO		GTP-binding nuclear prote	67.80029996	2	5.14	
mss-15052007-00003	<filename> (F1 25-26_5001)	\\spare\c_drive\b\IMA1_YEAST		Importin subunit alpha (Ka	64.13	2	2.77	
mss-15052007-00003	<filename> (F1 25-26_5001)	\\spare\c_drive\b\RS19A_YEAST		40S ribosomal protein S19	60.79	1	6.25	
mss-15052007-00003	<filename> (F1 25-26_5001)	\\spare\c_drive\b\EF1A_YEAST		Elongation factor 1-alpha (	54.44029996	3	5.68	
mss-15052007-00003	<filename> (F1 25-26_5001)	\\spare\c_drive\b\SYKC_YEAST		Lysyl-tRNA synthetase, cy	52.63726658	3	4.23	
mss-15052007-00003	<filename> (F1 25-26_5001)	\\spare\c_drive\b\XPO1_YEAST		Exportin-1 (Chromosome r	49.29029996	2	1.48	
mss-15052007-00004	<filename> (F2 26_5-28002)	\\spare\c_drive\b\ENO2_YEAST		Enolase 2 (EC 4.2.1.11) (2	64.63	2	4.58	
mss-15052007-00004	<filename> (F2 26_5-28002)	\\spare\c_drive\b\IMB4_YEAST		Importin beta-4 subunit (K	237.1729996	13	11.32	
mss-15052007-00004	<filename> (F2 26_5-28002)	\\spare\c_drive\b\GSP1_ASHGO		GTP-binding nuclear prote	192.6908999	4	9.81	
mss-15052007-00004	<filename> (F2 26_5-28002)	\\spare\c_drive\b\EF1A_YEAST		Elongation factor 1-alpha (	134.7014998	7	10.48	
mss-15052007-00004	<filename> (F2 26_5-28002)	\\spare\c_drive\b\XPO1_YEAST		Exportin-1 (Chromosome r	115.1381664	7	6.18	
mss-15052007-00004	<filename> (F2 26_5-28002)	\\spare\c_drive\b\SXM1_YEAST		Importin beta SXM1 (karyc	115.0208999	6	5.61	

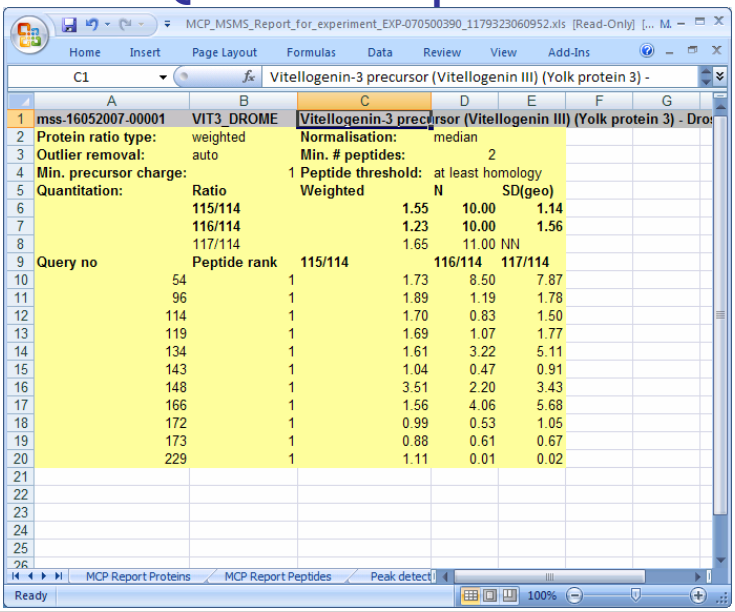
MCP Report Proteins | MCP Report Peptides | Peak detection parameters | False discovery rates

Ready

MASCOT : *MCP Guidelines without tears* 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



MCP_MSMS_Report_for_experiment_EXP-070500390_1179323060952.xls [Read-Only]

Home | Insert | Page Layout | Formulas | Data | Review | View | Add-Ins

C1: Vitellogenin-3 precursor (Vitellogenin III) (Yolk protein 3) -

	A	B	C	D	E	F	G
1	mss-16052007-00001	VIT3_DROME	Vitellogenin-3 precursor (Vitellogenin III) (Yolk protein 3) - Dro				
2	Protein ratio type:	weighted	Normalisation:	median			
3	Outlier removal:	auto	Min. # peptides:	2			
4	Min. precursor charge:	1	Peptide threshold:	at least homology			
5	Quantitation:	Ratio	Weighted	N	SD(geo)		
6		115/114	1.55	10.00	1.14		
7		116/114	1.23	10.00	1.56		
8		117/114	1.65	11.00	NN		
9	Query no	Peptide rank	115/114	116/114	117/114		
10	54	1	1.73	8.50	7.87		
11	96	1	1.89	1.19	1.78		
12	114	1	1.70	0.83	1.50		
13	119	1	1.69	1.07	1.77		
14	134	1	1.61	3.22	5.11		
15	143	1	1.04	0.47	0.91		
16	148	1	3.51	2.20	3.43		
17	166	1	1.56	4.06	5.68		
18	172	1	0.99	0.53	1.05		
19	173	1	0.88	0.61	0.67		
20	229	1	1.11	0.01	0.02		
21							
22							
23							
24							
25							
26							

MCP Report Proteins | MCP Report Peptides | Peak detect

Ready 100%

MASCOT : *MCP Guidelines without tears* MATRIX SCIENCE

Guideline 4 covers Quantitative proteomics results. The experiment we just carried out did not contain any quantitation results so these results of 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 the report these results and settings in the MCP publication report.

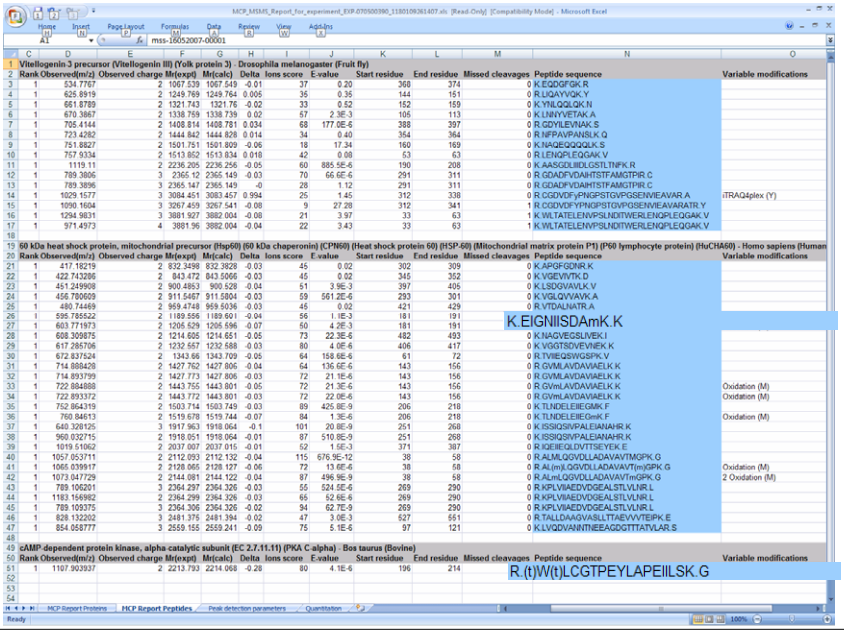
Molecular Cellular Proteomics publication guidelines

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

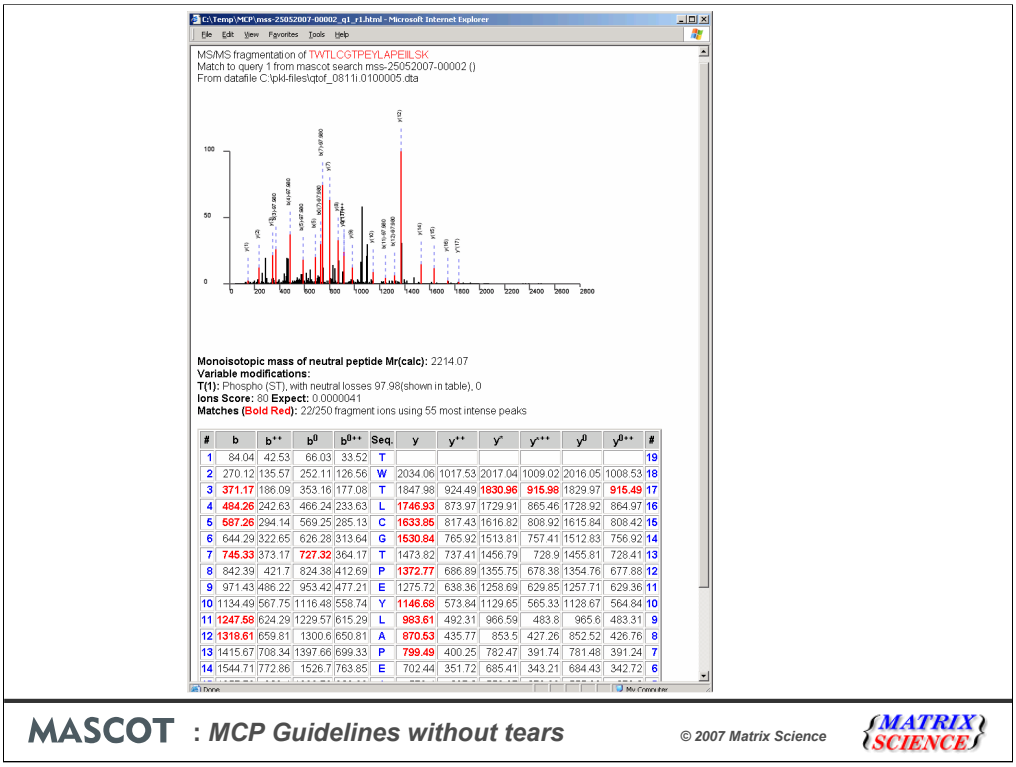
Guideline 5 is specific to modified peptides and requires that we supply additional information about post translationally modified peptides.

Guideline 5: Post-translational modifications



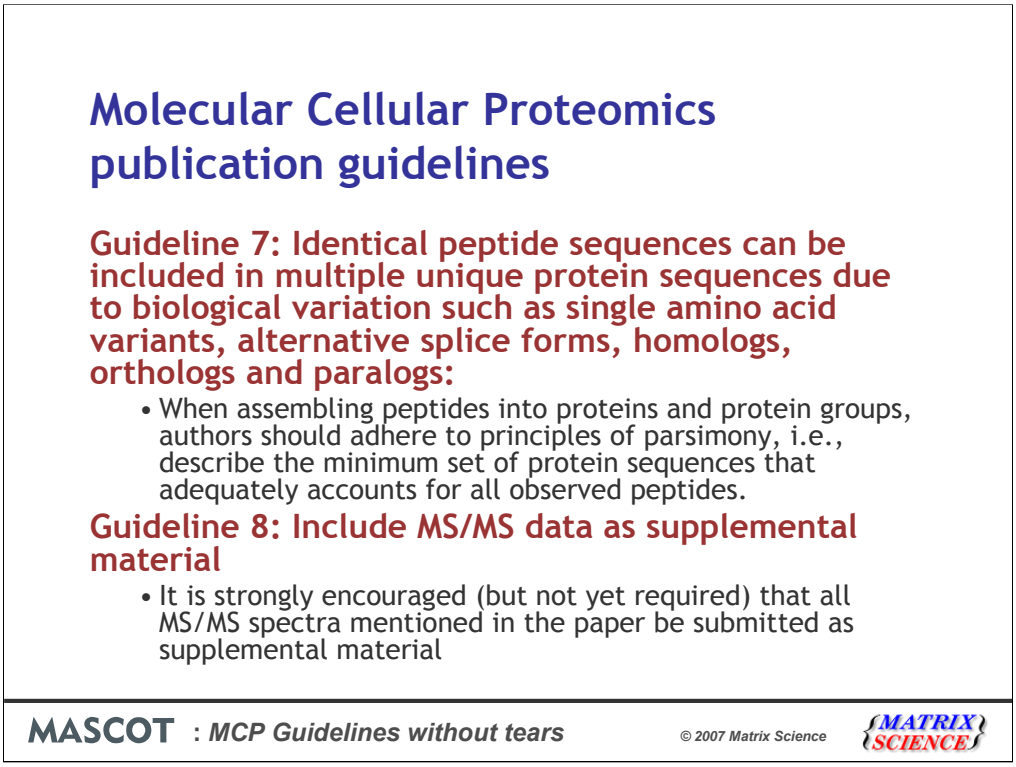
MASCOT : *MCP Guidelines without tears* MATRIX SCIENCE

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 ions 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.



#	b	b++	b0	b0++	Seq.	y	y++	y*	y*++	y0	y0++	#
1	84.04	42.53	66.03	33.52	T							19
2	270.12	135.57	252.11	126.56	W	2034.06	1017.53	2017.04	1009.02	2016.05	1008.53	18
3	371.17	186.09	353.16	177.08	T	1847.98	924.49	1830.96	915.98	1829.97	915.49	17
4	484.26	242.63	466.24	233.63	L	1746.93	873.97	1729.91	865.46	1728.92	864.97	16
5	587.26	294.14	569.25	285.13	C	1633.85	817.43	1616.82	808.92	1615.84	808.42	15
6	644.29	322.65	626.28	313.64	G	1530.84	765.92	1513.81	757.41	1512.83	756.92	14
7	745.33	373.17	727.32	364.17	T	1473.82	737.41	1456.79	728.9	1455.81	728.41	13
8	842.39	421.7	824.38	412.69	P	1372.77	686.89	1355.75	678.38	1354.76	677.88	12
9	971.43	486.22	953.42	477.21	E	1275.72	638.36	1258.69	629.85	1257.71	629.36	11
10	1134.49	567.75	1116.48	558.74	Y	1146.68	573.84	1129.65	565.33	1128.67	564.84	10
11	1247.58	624.29	1229.57	615.29	L	983.61	492.31	966.59	483.8	965.6	483.31	9
12	1318.61	659.81	1300.6	650.81	A	870.53	435.77	853.5	427.26	852.52	426.76	8
13	1415.67	708.34	1397.66	699.33	P	799.49	400.25	782.47	391.74	781.48	391.24	7
14	1544.71	772.86	1526.7	763.85	E	702.44	351.72	685.41	343.21	684.43	342.72	6

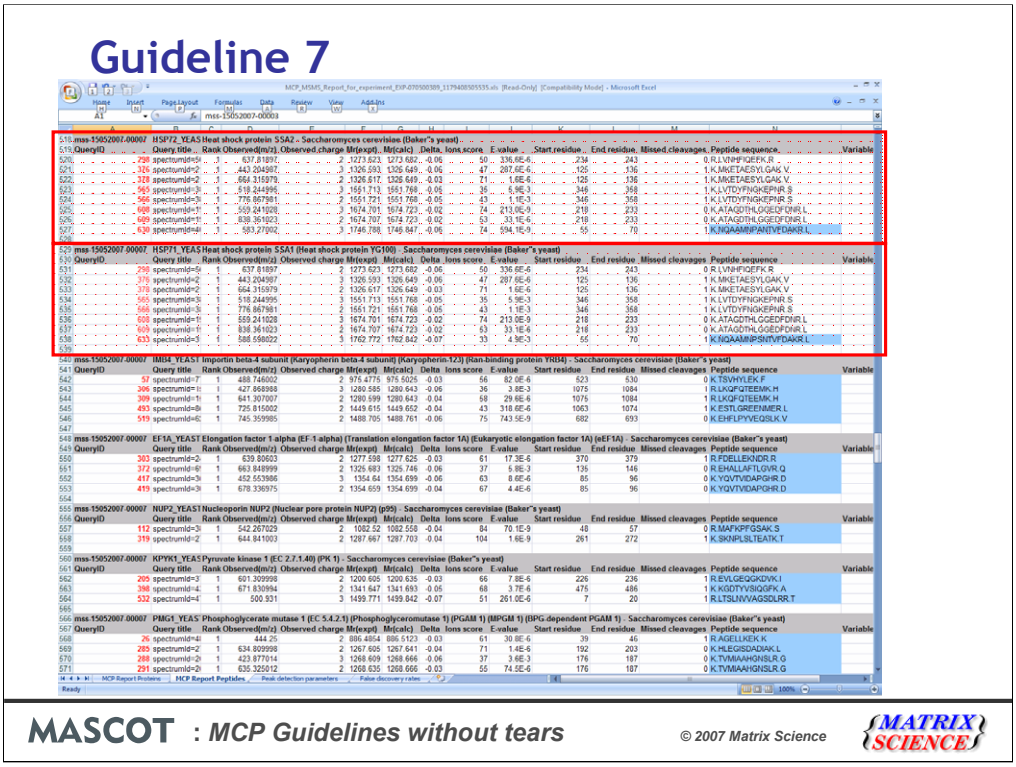
As for single peptide identifications, the MCP guidelines require annotated spectra for the modified peptides are submitted. These are included in the exported zip file in the html files.



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

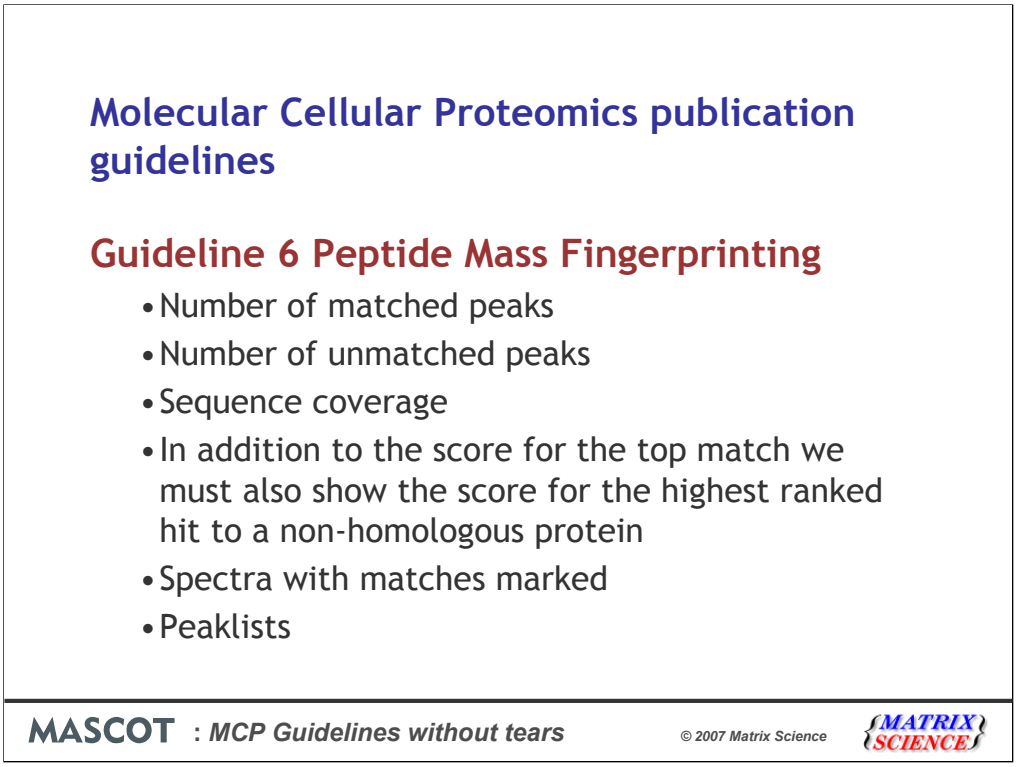
As mentioned earlier Peak lists produced during the automated data analysis with Mascot Daemon and saved as mzdata.xml files and are suitable for submitting as supplemental material if you choose to follow guideline 8.

Guideline 7 is concerned with which proteins are reported.



Because we've only exported approved protein hits for the MS/MS report, meeting the main requirement of guideline 7 (describing the minimum set of proteins that adequately accounts for all observed peptides) is handled by the user when they carryout protein hit approval. However, another requirement is that any peptides shared amongst multiple proteins and those unique to a specific protein should be clearly indicated. In the MCP MS/MS report exported from Mascot Integra this is achieved by highlighting peptides unique to a specific protein with a light blue background. If we take a look at these two protein hits – both heat shock proteins from budding yeast – we can easily see that the two different proteins are identified by 4 shared peptides and one unique peptide each.

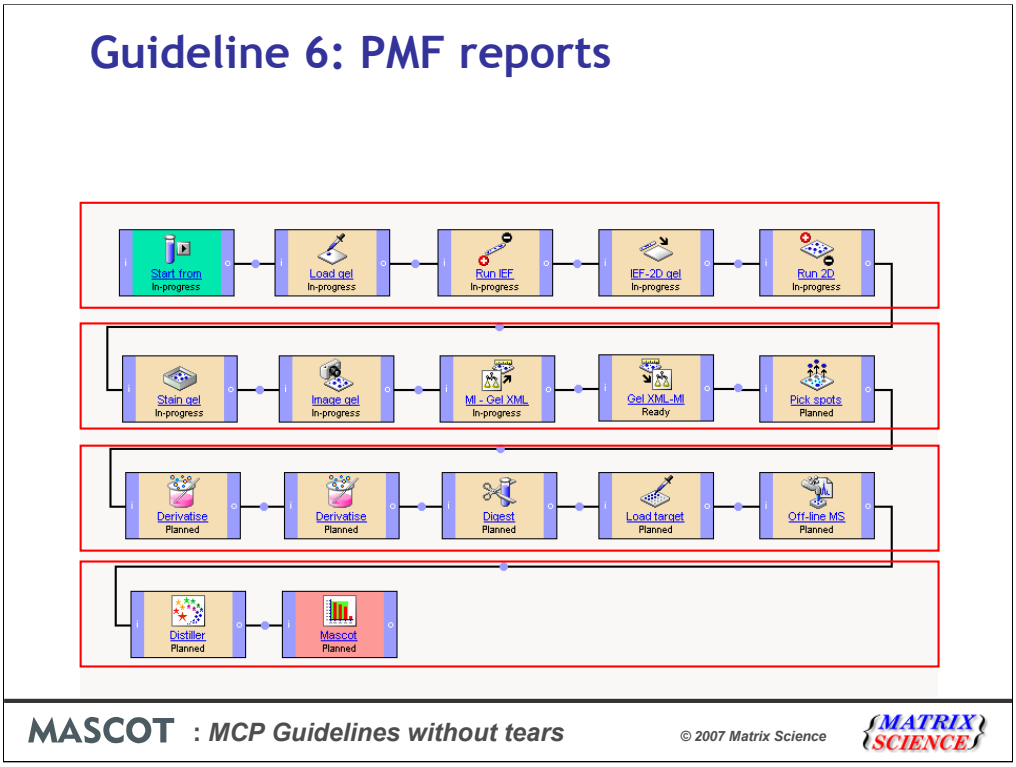
The other thing to look for here is the bold-red status of the QueryID, which matches the bold red settings on a standard mascot report (where bold is for the 1st time a query is used in the report, red for the top ranking match to the query).



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.



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:

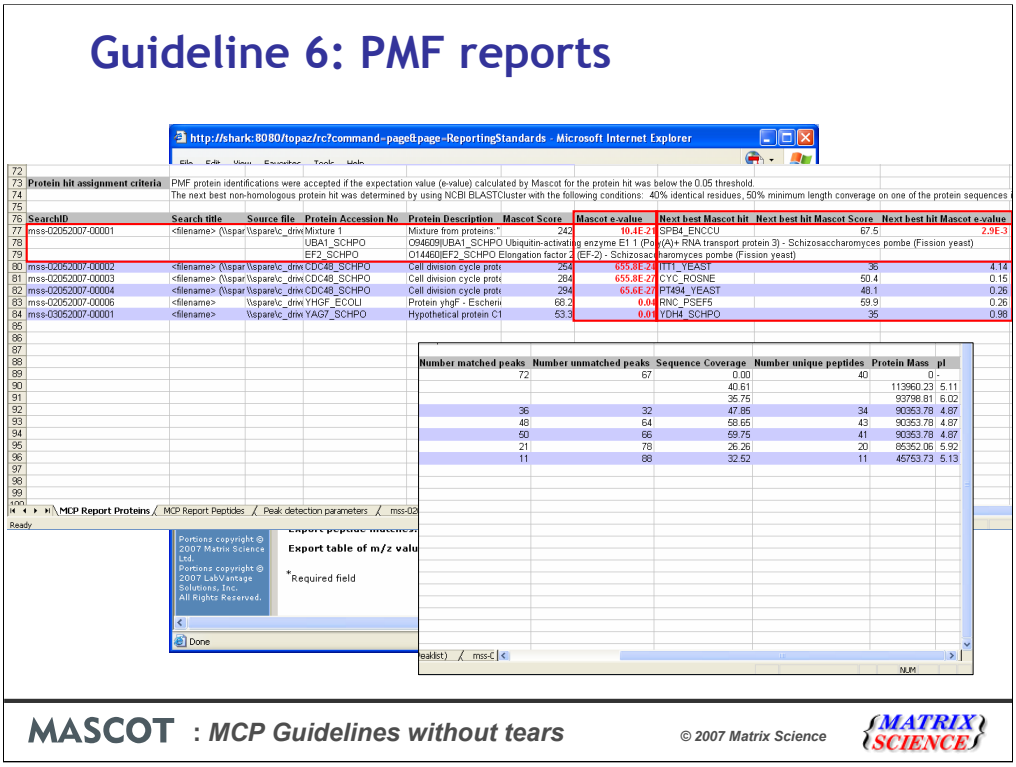
Preparation and running of the 2D gel;

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.



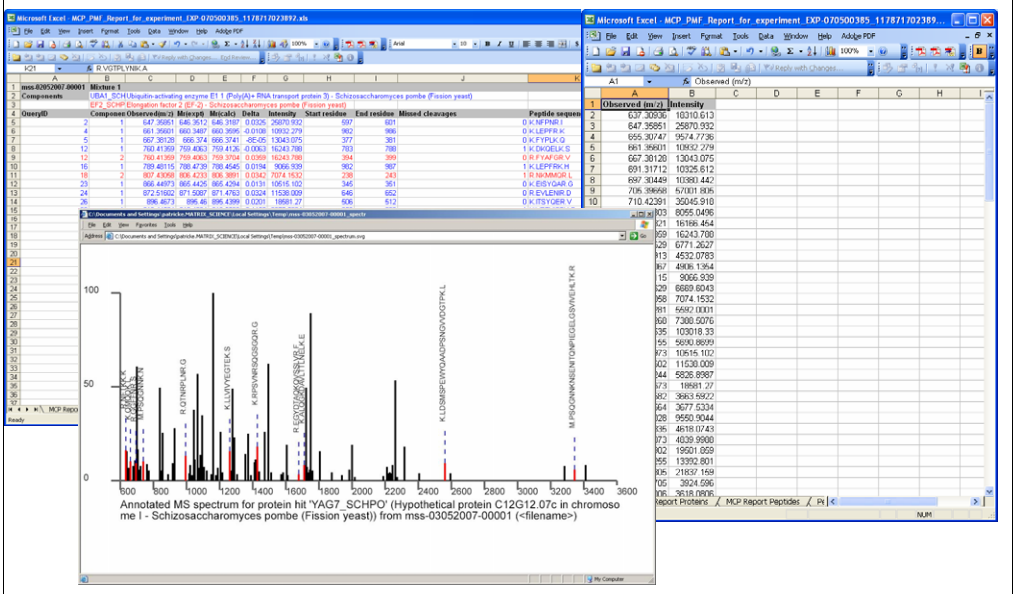
As for the MS/MS report the experiment and database searches are selected for exporting. One big difference is that unapproved search results with a protein hit e-value filter are used rather than approved search results. As for the MS/MS report, additional information has to be entered by the user before 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 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 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.

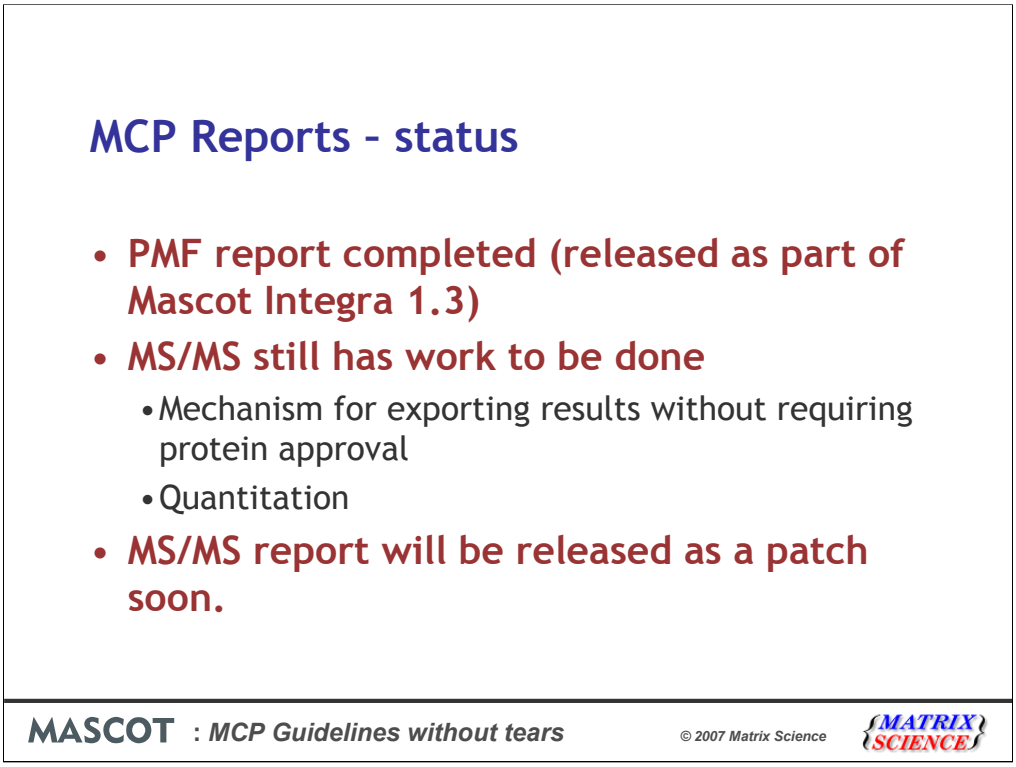
Guideline 6: PMF reports



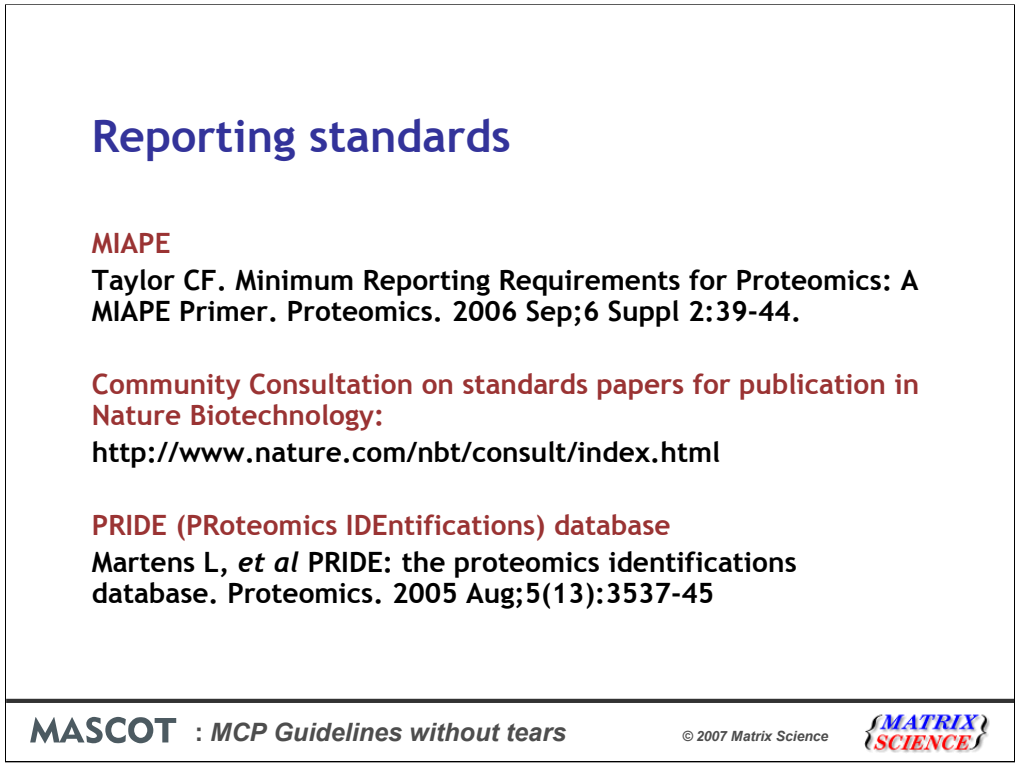
Individual peptide information is reported in a separate worksheet.

As are mass intensity lists for each of the searches.

Finally an image of the spectra with the matches marked on it is saved for each of the searches.



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

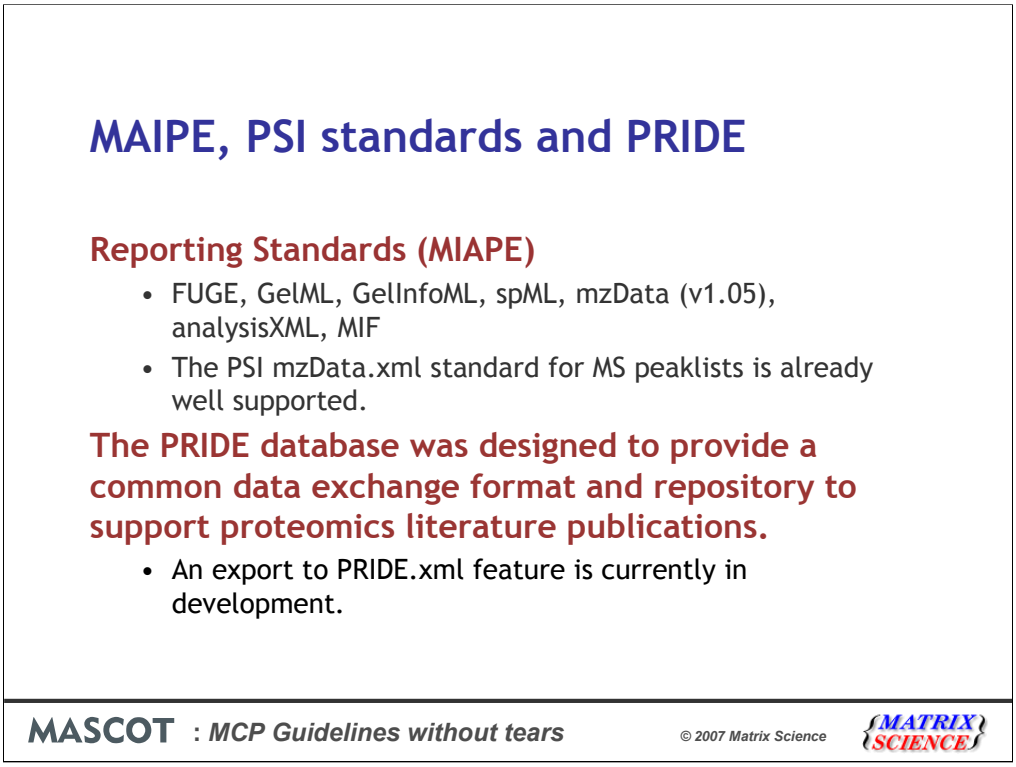


The MCP guidelines specify the types of information that should be reported but not the format for the information.

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 MAIPE 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.



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