

MASCOT*Integra*

Examples of real-life problem solving with Mascot Integra

MASCOT : Problem solving with Mascot Integra © 2006 Matrix Science



We have introduced Mascot Integra at previous ASMS user meetings so I thought that we would cover something different this time.

For those that are not familiar with Mascot Integra it is our solution to proteomics data management and you can find out more about it on our website or see it in action at our booth, number 39, during the conference.

This time I'm going to talk about solving real life problems with Mascot Integra.

Requirements

I need experiment condition tracking

I need sample tracking

I need flexible reporting

I need to analysis large data sets

What sort of requirements are we going to ask of Mascot Integra?

Firstly I need to track experiment conditions.

Secondly I want to know what happened to my samples three or more months later

Thirdly I want to create reports with the information that I'm interested in.

Finally I want to analysis large data sets and filter the identified proteins by different constraints

1. I need experiment condition tracking

Load the results in to the database as fast as possible

Track as many details as possible

An experiment can be as simple as this:



MASCOT : Problem solving with Mascot Integra © 2006 Matrix Science



The need for experiment condition tracking is split in to two camps:

Those that want to load the results into the database as fast as possible and go straight to analysis and reporting

And those that would like to track as many details as possible about the experiment from within the database

We can support both camps and anywhere in-between.

I going to show some examples of minimal and detailed experiments.

First the minimal experiment.

Experiments are modelled by linking together experimental tasks selected from the library

It can be a simple as this:

Start from a sample

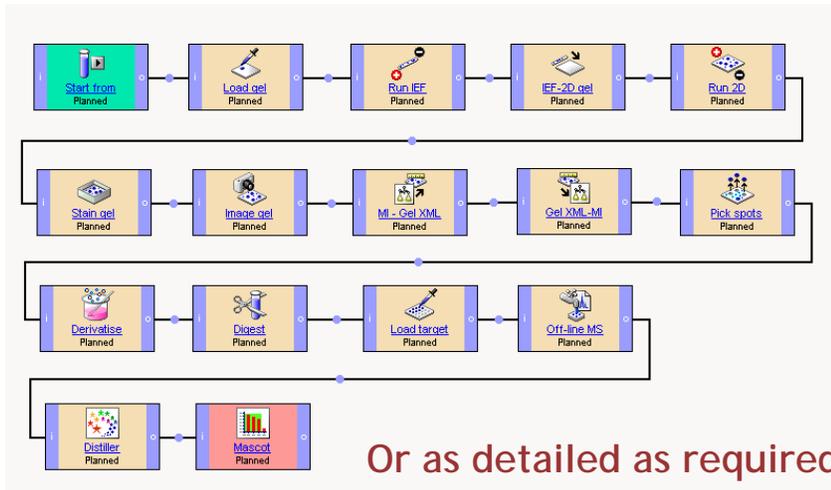
Import existing MS or MS/MS peaklists

And perform a Mascot search.

In this experiment we don't capture any additional information, parameters, or conditions about the experiment.

But it allows us to search one or multiple data files quickly and capture the results, and link the search results back to the original sample.

Experiment modelling



MASCOT : Problem solving with Mascot Integra

© 2006 Matrix Science

MATRIX
SCIENCE

This 2D gel protocol is an example of a more detailed experiment model.

I'm using many more different tasks to model this experiment.

Starting from our samples

we load and run the gels

Followed by staining and imaging

A link to the gel images is attached to the image gel task

Integra can interface with NonLinear Dynamics Progenesis and GE Healthcare DeCyder via xml files.

In this case Integra exports an xml file with information about the gels and their images that is imported into Progenesis

The gel images are then processed with Progenesis as per normal.

Once the gel spots have been selected for analysis we export an xml file from Progenesis.

This xml file is then imported back into Integra

and the spot list is processed generating a new gel spot sample for each selected spot.

Then we continue through the experiment reducing and alkylating the proteins before in-gel digestion with trypsin

The samples are then analysed by mass spectrometry

The raw data are processed by Mascot Distiller and Mascot Daemon submits the searches to the Mascot Server.

So, in this example, we will have captured all the experimental parameters and conditions from the processing of the sample to the MS analysis and data processing.

Experimental data capture

EXP-060200268-967 Digest In-progress

Data Set: Status Instrument: LabBook Ref:

Digestion Details Variant:1 Instance:1

Parameter	Type	Rep	Entered Value	Unit
<input type="checkbox"/> Protocol Id	Standard	1		
<input type="checkbox"/> Digestion buffer	Standard	1	25mM Ambic	
<input type="checkbox"/> Digestion buffer vol	Standard	1	50	
<input type="checkbox"/> Buffer volume units	Standard	1	ul	
<input type="checkbox"/> Incubation temp	Standard	1	37	C
<input type="checkbox"/> Incubation Time	Standard	1	4	
<input type="checkbox"/> Incubation time unit	Standard	1	hours	
<input type="checkbox"/> Storage type	Standard	1		
<input type="checkbox"/> Instrument Id	Standard	1		
<input type="checkbox"/> Stop buffer used	Standard	1		
<input type="checkbox"/> Stop buffer volume	Standard	1		
<input type="checkbox"/> Stop buffer units	Standard	1		

Data Set: Status Instrument: LabBook Ref:

Mascot Enzyme Variant:1 Instance:1

Parameter	Type	Rep	Entered Value	Unit
<input type="checkbox"/> Enzyme id	Standard	1	Trypsin	
<input type="checkbox"/> Enzyme Volume used	Standard	1	2	
<input type="checkbox"/> Enzyme volume units	Standard	1	ul	

Data Set: Status Instrument: LabBook Ref:

Sample volume used Variant:1 Instance:1

Parameter	Type	Rep	Entered Value	Unit
<input type="checkbox"/> Sample volume used	Standard	1	10	
<input type="checkbox"/> Sample volume units	Standard	1	mg	

MASCOT : Problem solving with Mascot Integra

© 2006 Matrix Science



Most tasks have a data entry page. Here is a view of the data entry page for an enzyme digest task.

The yellow boxes indicate values that are required to perform the task.

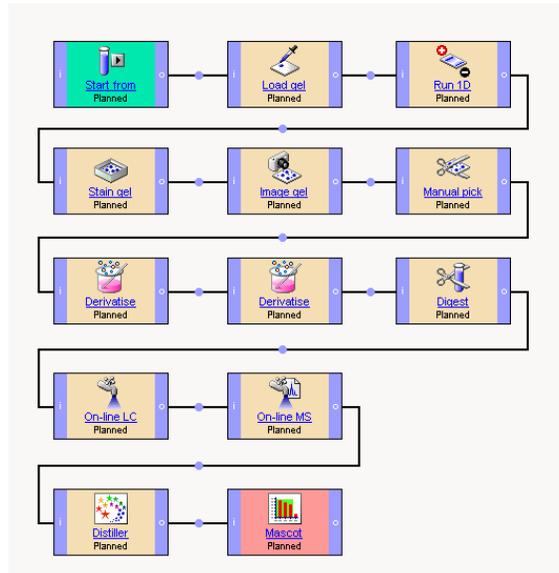
The other boxes are optional, but the more data that is entered, the more scope there will be for data mining of the experimental results.

Single tasks through to complete experiment plans can be preconfigured and saved as templates.

So, for example, if you perform all your digestions at 37C for 4 hours then these values can be predefined so that you don't have to enter them each time you run the task.

Some values are selected from a controlled, user defined list. For this task, it's the digestion and stop buffers and digestion robot or instrument.

Experiment modelling



MASCOT : Problem solving with Mascot Integra

© 2006 Matrix Science

MATRIX
SCIENCE

Here is a third example of a detailed model of a protein mixture separated on a 1D gel.

The sample loaded and separated on a 1D gel

The gel is stained and the protein bands are excised

Followed by in gel digestion

LCMSMS

Peak picking and mascot database searching.

Again we have captured all the experimental parameters and conditions from the processing of the sample to the MS analysis and data processing

Instrument and gel package integration

Instrument integration via sample sheet exchange (Excel, CSV, tab delineated files)

Gel package integration via XML import and export

Currently supported

- GE Healthcare DeCyder
- NonLinear Dynamics Phoretix

Adding in the next release

- Bio-Rad PDQuest

MASCOT

: Problem solving with Mascot Integra

© 2006 Matrix Science

MATRIX
SCIENCE

Rather than trying to run the instrumentation directly from Mascot Integra we have implemented a samplesheet exchange system. During the relevant step of running an experiment, you produce an Excel or CSV samplesheet from within the Integra system which you can then load into the Instrument datasystem. Nearly all instrument datasystems will import one or more of the supported formats (Excel, CSV and tab delineated).

This system is highly flexible. You can design your own samplesheets for use within the system. So if Thermo add a new column to the Xcalibur samplesheet then you can add this yourself into the system – you don't have to wait for us to add the column. Likewise if you bought a spotting robot which we had not designed a samplesheet for then you could simply add a new design into the system for it.

We also integrate with a number of Gel Packages. This is carried out using XML exchange between Mascot Integra and the Gel package software. We currently support the GE Healthcare DeCyder software and the NonLinear Dynamics family of 2D Gel analysis software and will be adding support for the Bio-Rad PDQuest package in the next release.

2. I need sample tracking

Relationships between samples are tracked at each and every experimental step

These can be 1-1, 1-many, many-1 or many-many relationships

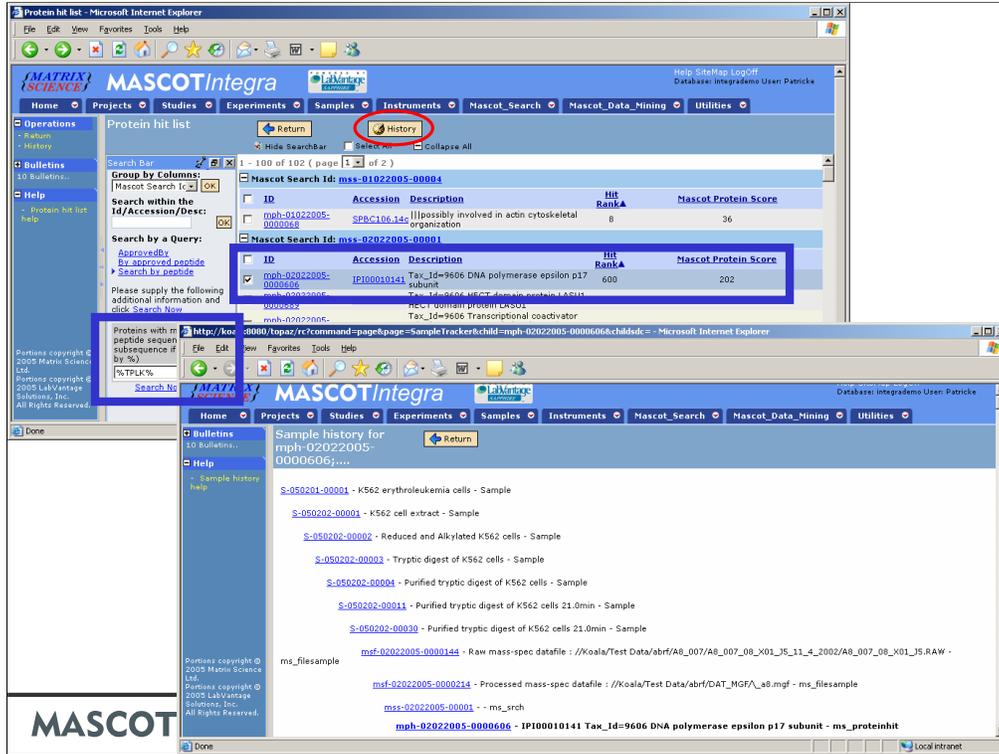
Allows you to easily answer the question - where did this search result come from (or the other way round)?

Now on to the second requirement of sample tracking.

One of the key advantages of tracking your experiments and results through Mascot Integra is that at each step of the experiment the relationships between samples are tracked and stored, readily accessible, in the database.

These relationships can be 1-1 (i.e. doing a simple digest to produce a child sample), 1-many (i.e. splitting a sample into multiple aliquots), many-1 (i.e. creating a single mixture from multiple starting samples) or many-many (i.e. creating multiple mixtures from multiple starting samples).

Because we have tracked all of these relationships, then answering the question 'where did this search result come from' 12 months down the line is now trivial.



Here I've done a query against our Mascot Integra system and retrieved a list of all Mascot protein hits which have a peptide match that contains the subsequence TPLK. Then to see were any one of those protein hits originates from all I have to do is select the protein hit and click on the History button.

We can see that all the intermediate steps from the protein hit back to the original cell sample.

3. I need flexible reporting

Excel reporting

- Analysis of all the peptides identified by MS/MS from a single or multiple analysis
- Produce a mass and retention time overview
- Protein vs DNA (human genome) data comparison

Advanced reporting

- Plate assay report
- PMF report that satisfies MCP guide lines

MASCOT

: Problem solving with Mascot Integra

© 2006 Matrix Science

MATRIX
SCIENCE

Once that I have samples and data in the system I would like to perform some analysis and create some reports.

Mascot Integra uses Excel for custom reporting.

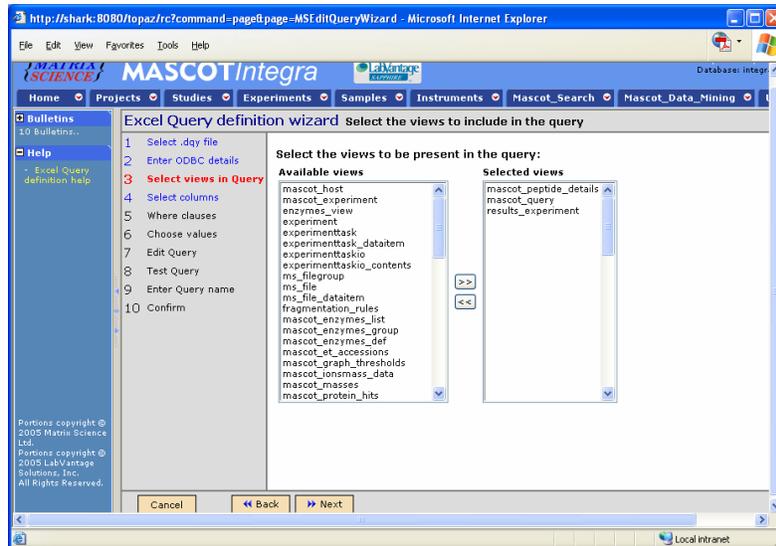
For more complex data mining, involving multiple SQL statements and further data processing, the Integra database can be mined using custom scripts or programs.

Pretty much any programming language can be used, for example Perl, Java, Visual Basic or C++. You could even use Visual Basic macros in an Excel sheet.

I'm going to show how an excel report is designed and three examples from the currently available report set.

Then I'm going to show some examples of what you might want to do with more advanced reporting.

Excel query - choose the views



MASCOT : Problem solving with Mascot Integra

© 2006 Matrix Science



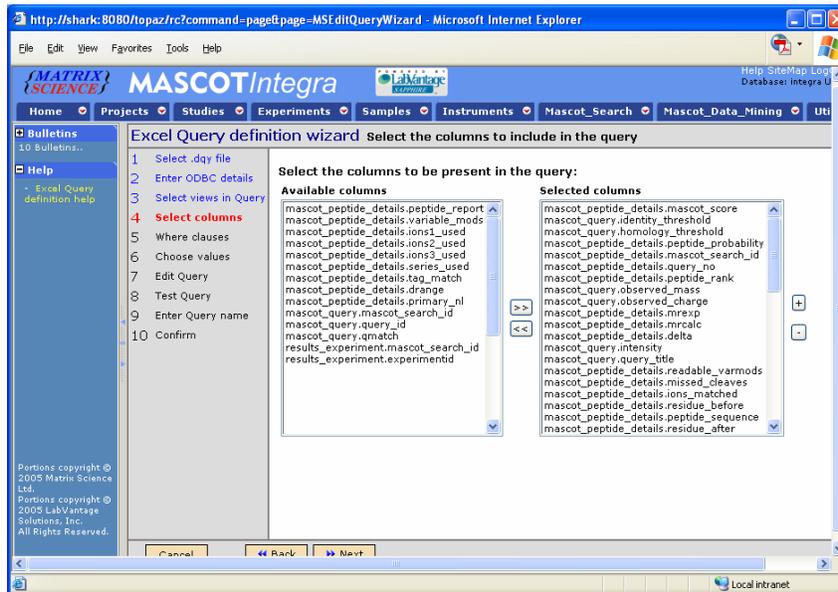
The first excel report is on the properties of the peptides identified from a MuDPIT run.

All the available views are listed in the left panel

from those I select the views related to peptides identified in a search.

The relationships between the different views and the column names for each view are documented in the online help.

Excel query - select the columns



MASCOT : Problem solving with Mascot Integra

© 2006 Matrix Science



Next we select the columns to be reported

The available columns from our chosen views are displayed in the left panel.

And the selected columns are shown in the right panel.

You can order the columns as you prefer.

Excel query - search by experiment id and sort by peptide mascot score

The screenshot shows the 'Excel Query definition wizard' in Microsoft Internet Explorer. The browser address bar shows 'BD/topaz/irc?command=page&page=MSFEditQueryWizard'. The page title is 'MASCOT Integra'. The navigation menu includes 'Projects', 'Studies', 'Experiments', 'Samples', 'Instruments', 'Mascot_Search', 'Mascot_Data_Mining', and 'Utilities'. The wizard is currently on step 5, 'Where clauses'. The instructions state: 'Specify clauses to limit the data returned by the query, or to control the order of the query. Specify a value of ? if you wish the clause value to be determined at runtime. Please not that you cannot specify runtime parameters for 'LIKE' clauses.' The table below shows two clauses:

Clause type	Column name	Operator	Value	
Where	results_experiment_v.experimentid	Equals	?	Remove Clause
Order By	mascot_peptide_details_v.mascot_score	Descending	?	Remove Clause

At the bottom of the wizard, there are 'Cancel', 'Back', and 'Next' buttons. The footer of the page contains the 'MASCOT' logo, the text ': Problem solving with Mascot Integra', the copyright notice '© 2006 Matrix Science', and the 'MATRIX SCIENCE' logo.

Next we specify clauses to limit the peptides to an experiment ID which will be defined at run time

And a second clause to sort the peptides by mascot score.

Mascot Integra builds the SQL statement which can then be edited (if required), tested and saved.

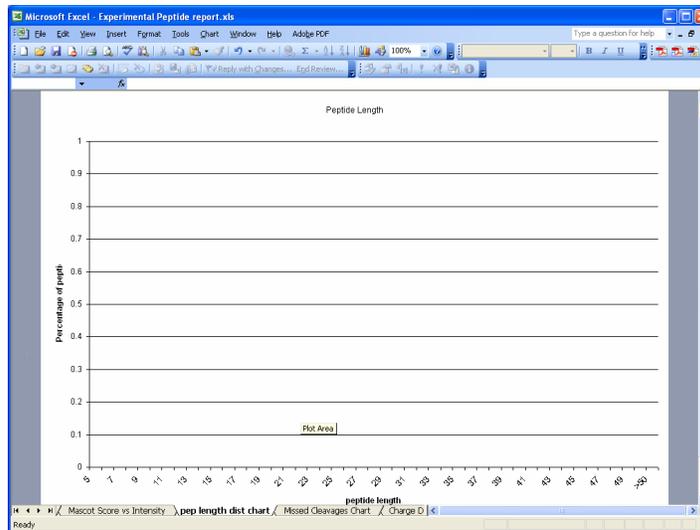
Build an Excel template

MASCOT : Problem solving with Mascot Integra © 2006 Matrix Science **MATRIX SCIENCE**

Next I open excel and build a template for the results.

I specify where the data will be placed within the worksheet and can program additional calculations on the data and design graphs.

Build an Excel template



MASCOT : Problem solving with Mascot Integra

© 2006 Matrix Science



I can add additional graphs and calculations to the template at a later date as long as the query stays the same.

The excel sheet is then saved on the Mascot Integra server and added to the Excel report templates. Other users can then run the same report on their own searches/samples without having to know anything about the underlying SQL.

Test sample

Cleavable ICAT analysis of rat brain sample
SCX separation into fractions and analysed by
LCMS/MS on a LTQ-FT

Automated analysis performed from Mascot Integra
with peak picking by Mascot Distiller and database
searching against NCBI nr with Mascot server.

In total there were 97,353 queries.

Resulting in 9257 peptide hits above the Mascot
homology or significance score.

Data generously provided by Pascal Wather and
Bertran Gerrits

MASCOT : Problem solving with Mascot Integra © 2006 Matrix Science



I had previously processed a data from large ICAT analysis of the rat brain proteome that had been separated by strong cation ion exchange followed by LCMS/MS on a LTQ-FT.

The data analysis was performed automatically with Mascot Distiller and Mascot Server.

The search results were then available for reporting from Mascot Integra.

Excel query - export of custom reports

ms_XLSReport 'xls-00020' was successfully deleted.

ID	Description
<input type="radio"/> xls-00001	All significant protein hits from an experiment (only valid for PMF experiments)
<input type="radio"/> xls-00005	All approved protein hits for an experiment
<input type="radio"/> xls-00006	Mascot searches containing specified accession
<input type="radio"/> xls-00007	Project Details
<input type="radio"/> xls-00008	Mascot searches within a specified study
<input type="radio"/> xls-00009	Protein hits from an experiment task with an evaluate below X
<input type="radio"/> xls-00010	Gel and gelspot details
<input type="radio"/> xls-00012	Protein hit details for a specified Mascot search
<input type="radio"/> xls-00013	Protein hit details by id
<input type="radio"/> xls-00014	Protein hits from gel lane in 1D GelCMS experiment
<input type="radio"/> xls-00015	Comparison report: No accession or shared sig peptide match in Control search
<input type="radio"/> xls-00016	Comparison report: No accession and no sig peptide has match in Control
<input type="radio"/> xls-00017	Comparison report: No accession or peptide match in the Control
<input checked="" type="radio"/> xls-00018	Experiment peptide report MS/MS
<input type="radio"/> xls-00019	Mass vs Retention Time vs Intensity (MASCOT ARRAY)

MASCOT : Problem solving with Mascot Integra

© 2006 Matrix Science



Lets export all the peptides above the homology score that have been identified by ms/ms from the experiment.

First we select the report template

Excel query - export of custom reports

The screenshot shows the Mascot Integra web interface. At the top, there are navigation tabs: Projects, Studies, Experiments, Samples, Instruments, Mascot_Search, Mascot_Data_Mining, and Utilities. Below the tabs, there is a 'Download Excel Report' section with a text input field for 'Enter value for sheet 1: cell B-1 (experiment_id)' containing 'EXP-060400267' and a 'NEXT >>' button. A 'Select Experiment - Microsoft Internet Explorer' dialog box is open, displaying a list of experiments. The dialog has a 'Cancel' button and a search field. The list is organized into studies:

ID	Description
EXP-060300245	Mascot search import
EXP-060300246	Error tolerant searches from peaklist
Study S-060300033 96Well plate BI test	
EXP-060300247	Example BI plate with data
EXP-060300248	plate 2
EXP-060300249	BI plate 3
EXP-060300250	plate 4
EXP-060300251	BI test full plate 5
EXP-060300254	BI test full plate 6
EXP-060300255	plate 7 Biphospho
EXP-060300256	plate 8 colour coding test
EXP-060300257	PMF plate
Study S-060400034 Minimal experiments	
EXP-060400260	Peak list example
EXP-060400261	Raw data example
EXP-060400262	Mascot Results example
Study S-060400035 2d gel	
EXP-060400259	2D gel analysis with Phoretix and MS task
Study S-060400036 ICAT wash through	
EXP-060400263	ICAT wash through example
Study S-060400037 COFRADIC partial model	
EXP-060400264	COFRADIC example three exps

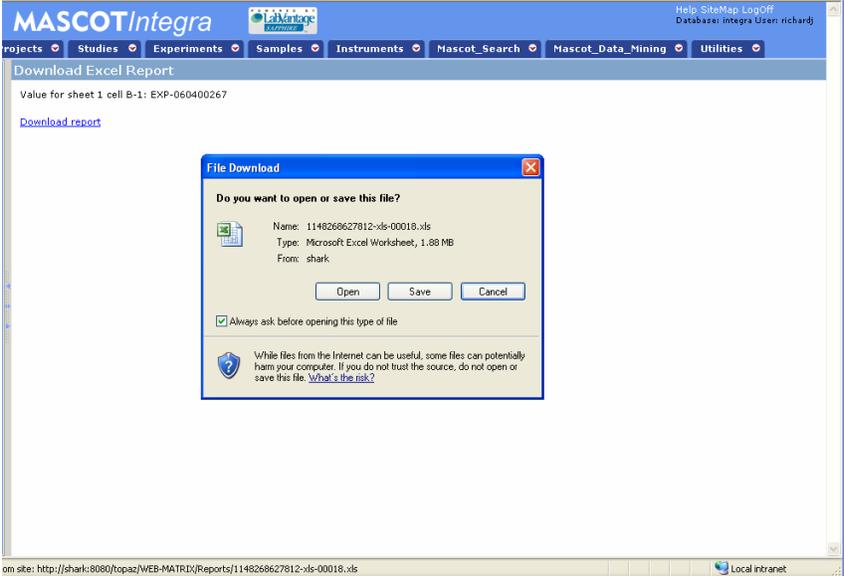
MASCOT : Problem solving with Mascot Integra

© 2006 Matrix Science



Then the mascot search id

Excel query - export of custom reports



The screenshot displays the MascotIntegra web interface. The main content area shows a 'Download Excel Report' section with the text 'Value for sheet 1 cell B-1: EXP-060400267' and a 'Download report' link. A 'File Download' dialog box is open, asking 'Do you want to open or save this file?'. The dialog provides the following information: Name: 1148268627812-xls-00018.xls, Type: Microsoft Excel Worksheet, 1.88 MB, and From: shark. It includes 'Open', 'Save', and 'Cancel' buttons, and a checked option for 'Always ask before opening this type of file'. A security warning at the bottom of the dialog states: 'While files from the Internet can be useful, some files can potentially harm your computer. If you do not trust the source, do not open or save this file. What's the risk?'. The browser's address bar shows the URL: http://shark:8080/topaz/WEB-MATRIX/Reports/1148268627812-xls-00018.xls. The footer of the page contains the Mascot logo, the text ': Problem solving with Mascot Integra', the copyright notice '© 2006 Matrix Science', and the Matrix Science logo.

And finally save the report as an Excel document to the hard drive.
The report can then be opened and edited in Excel.

Excel query - view the results

MASCOT_SCORE	IDENTITY_THRESHOLD	HOMOLOGY_THRESHOLD	PEPTIDE_PROBABILITY	MASCOT_SEARCH_ID	QUERY_NO	PEPTIDE_RANK	OBSERVED_MASS	OBSERVED_CHARGE	MREXP	MRCLAC	DELTA	INTENSITY	QUERY_TITLE
77.51	35	30	3.35499E-06	msa-26042006-00002	1948	1	1506.8238	1	1.624.815524	1.624.800598	0.014526	1278757.6	1147_Scan_2420 (m/z=2)
77.5	35	30	3.63638E-06	msa-26042006-00001	1716	1	1506.8213	1	1.624.814024	1.624.800598	0.014526	372567.30	1112_Scan_2455 (m/z=2)
77.5	35	23	3.39525E-06	msa-26042006-00003	1688	1	1495.7925	1	1.495.792524	1.495.770401	0.014523	1157134.1	1320_Scan_21 (m/z=2)
74.8	35	24	6.32295E-06	msa-26042006-00002	1938	1	1486.7923	1	1.486.792324	1.486.770401	0.014523	3617879.1148	Sum of 2 scans
39.08	35	27	0.022040614	msa-26042006-00004	1232	1	1074.6136	1	1.073.603264	1.073.589999	0.007255	1293449.9	1219_Scan_2654 (m/z=2)
38.84	29	19	0.005753683	msa-27042006-00002	1247	1	1320.7841	1	1.319.778624	1.319.768173	0.006861	489876.1508	Scan_2810 (m/z=2)
38.03	36	26	0.02627074	msa-26042006-01194	1305	1	1160.6987	1	1.159.691424	1.159.670066	0.006359	133481.3	1474_Scan_2512 (m/z=2)
38.34	33	26	0.01539968	msa-26042006-00002	736	1	909.4483	1	0.909.386664	0.909.370994	0.006066	749676.86146	Scan_1931 (m/z=19)
38.32	35	17	0.02770892	msa-26042006-00001	1188	1	1198.6342	1	1.197.628924	1.197.617699	0.009355	295938.4	1080_Scan_2408 (m/z=2)
38.01	35	16	0.0000279	msa-27042006-00003	1152	1	1168.634	1	1.168.634224	1.168.617699	0.009155	147439.75	1332_Scan_2448 (m/z=2)
37.95	36	25	0.03907109	msa-26042006-00001	1341	1	1302.6917	1	1.301.684424	1.301.676117	0.006307	4242650.9	979_Sum of 2 scans
37.63	36	20	0.04018942	msa-26042006-00002	1933	1	1302.6917	1	1.301.689824	1.301.676117	0.006707	480233.9	1072_Scan_21 (m/z=2)
37.79	36	17	0.041202731	msa-26042006-00003	1697	1	1302.6916	1	1.301.689824	1.301.676117	0.012107	74393.38	1188_Scan_2388 (m/z=2)
37.72	36	20	0.04120668	msa-27042006-00003	1289	1	1302.6914	1	1.301.684124	1.301.676117	0.009007	430447.38	1198_Scan_2633 (m/z=2)
37.28	35	25	0.031695175	msa-26042006-00003	1339	1	1169.6348	1	1.168.628724	1.168.620966	0.007369	249683.93	1246_Scan_2469 (m/z=2)
37.16	35	31	0.033259871	msa-26042006-00001	1130	1	1169.6349	1	1.168.627624	1.168.620966	0.006739	427363.84	1057_Scan_2776 (m/z=2)
36.51	32	30	0.023995706	msa-26042006-00004	1300	1	1117.4974	1	1.116.491224	1.116.490998	0.003198	134100.05	712_Scan_1815 (m/z=19)
36.5	33	19	0.035173419	msa-27042006-00003	1898	1	1311.7549	1	1.310.747624	1.310.737976	0.006648	96279.68	1526_Scan_2638 (m/z=2)
36.47	36	23	0.056599726	msa-26042006-00004	1363	1	1180.5933	1	1.159.598024	1.159.570066	0.009659	173876.55	1203_Scan_2628 (m/z=2)
34.84	35	25	0.060226667	msa-26042006-00003	1395	1	1180.5933	1	1.159.627024	1.159.619666	0.006455	104624.5	1197_Scan_2584 (m/z=2)
34.83	36	27	0.078135147	msa-26042006-00001	1117	1	1160.6098	1	1.159.598224	1.159.590966	0.007856	504521.66	1080_Scan_2382 (m/z=2)
33.91	35	23	0.07039918	msa-27042006-00003	1307	1	1311.7228	1	1.310.750524	1.310.740114	0.006211	303446.88	1197_Scan_2632 (m/z=2)
33.74	34	22	0.0006121	msa-26042006-00003	1200	1	1103.6386	1	1.102.631324	1.102.620934	0.01049	416995.6	1646_Sum of 2 scans
33.73	33	26	0.05189986	msa-27042006-00002	1232	1	1311.754	1	1.310.746724	1.310.737976	0.009748	1130411.6	1508_Scan_21 (m/z=2)
32.43	32	15	0.040491344	msa-26042006-00002	1149	1	1113.6693	1	1.111.662024	1.111.651031	0.010893	780760.3	1063_Scan_21 (m/z=2)
32.06	27	0	0.01697972	msa-26042006-00003	1809	1	1415.6173	1	1.414.610024	1.414.603424	0.0066	241827.73	953_Scan_1844 (m/z=19)
32.06	32	15	0.040491344	msa-26042006-00002	1149	1	1113.6693	1	1.111.662024	1.111.651031	0.010893	780760.3	1063_Scan_21 (m/z=2)
32.06	28	0	0.000624684	msa-26042006-00003	1635	1	1320.7882	1	1.319.789024	1.319.788173	0.012151	435768.22	1517_Scan_2872 (m/z=2)
32.23	34	27	0.091168007	msa-26042006-00002	998	1	1023.5321	1	1.022.524624	1.022.521886	0.002968	1412960.9	928_Scan_1968 (m/z=3)
32.17	37	16	0.19569535	msa-26042006-00003	1818	1	1415.6188	1	1.417.717224	1.417.698303	0.013421	88278.5	1181_Scan_2300 (m/z=2)
32.05	27	16	0.019647647	msa-26042006-00002	1743	1	1415.6188	1	1.414.639524	1.414.633424	0.0061	412119.44	726_Scan_1763 (m/z=19)
31.86	36	19	0.131613206	msa-26042006-00003	1421	1	1207.6618	1	1.206.654524	1.206.647798	0.009758	1584925.5	1207_Scan_2424 (m/z=2)
31.97	34	25	0.09625636	msa-26042006-00002	951	1	1023.5322	1	1.022.529024	1.022.521886	0.004608	801377.63	803_Scan_1978 (m/z=19)
31.95	35	26	0.11570996	msa-27042006-00003	1177	1	1213.623	1	1.212.615724	1.212.610716	0.005006	33201.29	1071_Scan_2448 (m/z=2)
31.95	35	0	0.03333917	msa-26042006-01195	249	1	797.32195	1	0.796.34574	0.796.316991	0.002107	9721.894	856_Scan_1521 (m/z=16)
31.16	36	18	0.18948028	msa-26042006-00001	1541	1	1417.7117	1	1.417.709724	1.417.698303	0.011421	476993.4	996_Scan_2284 (m/z=23)
30.88	32	20	0.001610883	msa-26042006-00001	1257	1	1234.5797	1	1.233.568424	1.233.559993	0.002537	139338.33	716_Scan_1938 (m/z=19)
30.47	32	19	0.076552047	msa-26042006-01194	253	1	1132.5289	1	1.131.521624	1.131.519623	0.003001	123486.46	1318_Scan_2531 (m/z=2)
30.18	30	15	0.059914659	msa-26042006-01194	995	2	976.44817	1	0.975.448994	0.975.440364	0.000207	86081.5	1112_Scan_2711 (m/z=2)
30.18	30	15	0.059914659	msa-26042006-01194	995	1	976.44817	1	0.975.448994	0.975.440364	0.000207	86081.5	1112_Scan_2711 (m/z=2)
29.48	36	21	0.2671701	msa-26042006-00003	1635	1	1316.7195	1	1.315.712224	1.315.7006	0.011724	1340122.8	1141_Scan_2086 (m/z=2)
29.37	34	17	0.00030663	msa-26042006-00001	1389	1	1243.6388	1	1.242.598724	1.242.590966	0.003539	1150203.2	1177_Scan_1939 (m/z=19)
29.05	35	29	0.205469302	msa-26042006-00002	1064	1	1074.6136	1	1.073.605524	1.073.589999	0.006235	1128414.1	1226_Sum of 2 scans
28.78	35	24	0.216073065	msa-26042006-01194	1175	1	1074.6084	1	1.073.601124	1.073.589999	0.002125	490239.97	1488_Scan_2816 (m/z=2)
27.88	32	14	0.130425147	msa-26042006-00004	1263	1	1113.6694	1	1.111.662124	1.111.651031	0.011063	484394.1633	Scan_2306 (m/z=2)
27.76	26	0	0.03810495	msa-27042006-00002	231	1	787.31982	1	0.786.32544	0.786.316891	0.004137	8402.1789	803_Scan_1968 (m/z=17)
27.65	35	21	0.016804707	msa-26042006-00001	1310	1	1277.752	1	1.276.748724	1.276.740006	0.006116	4442179.6	1198_Scan_21 (m/z=2)
27.63	35	27	0.01690172	msa-26042006-00002	787	1	959.50093	1	0.958.493664	0.958.49097	0.003684	538349.773	Scan_1934 (m/z=19)
27.59	35	24	0.13210424	msa-26042006-00002	1315	1	1198.6344	1	1.197.627124	1.197.617699	0.009555	1441974.1	1006_Scan_2200 (m/z=2)
27.43	36	25	0.470489244	msa-26042006-00002	705	1	930.5197	1	0.928.626264	0.928.620397	0.005697	83070.94	839_Scan_1931 (m/z=19)

MASCOT : Problem solving with Mascot Integra © 2006 Matrix Science

That 5 minutes of work designing a query produced a table with information on all the peptides identified in an experiment. In this case data on 9257 peptides.

Excel query

-what shall we do with the data

Observed mass vs. delta error

Peptide length

Missed cleavages distribution

Charge state distribution

AA frequency

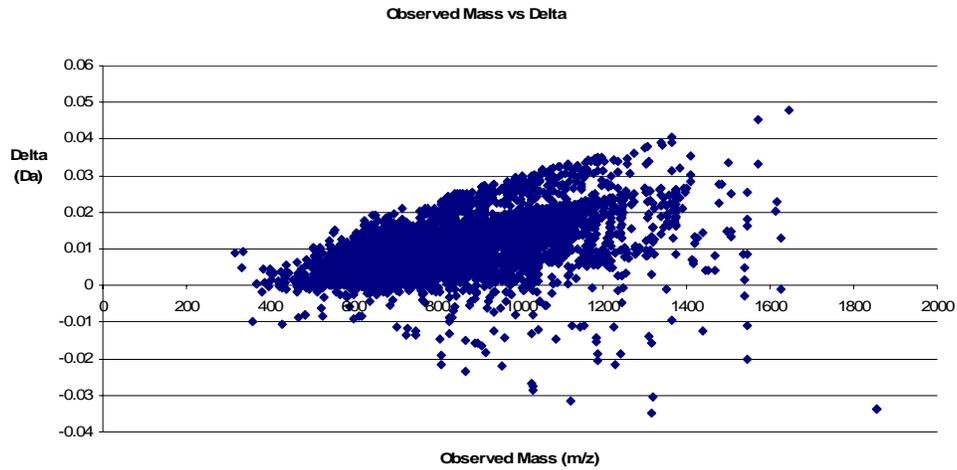
Mascot score vs. delta error

Mascot score vs. number of fragment ions

Most of you will be familiar with Excel and already have enough knowledge to generate sophisticated calculations and graphs.

Here are some of the calculations and graphs that I built.

Observed mass vs delta error



MASCOT : Problem solving with Mascot Integra

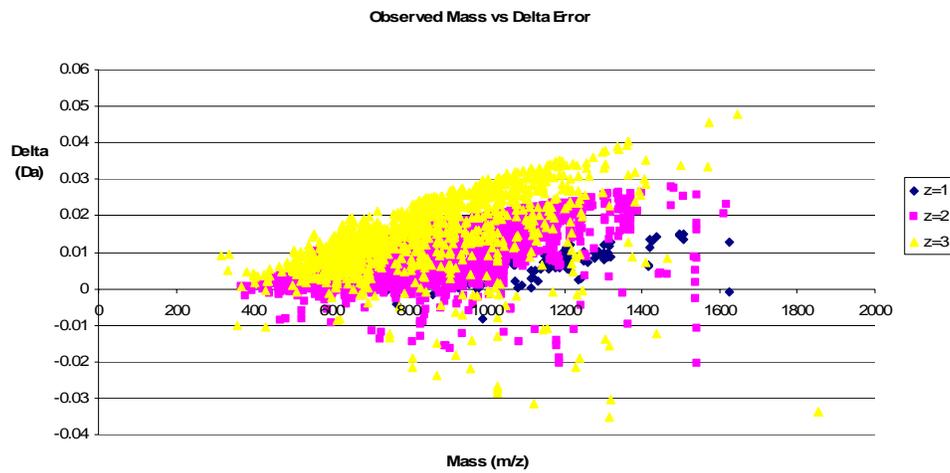
© 2006 Matrix Science

MATRIX
SCIENCE

Looking at the observed mass vs delta error graph there seemed to be multiple populations probably relating to the charge state.

I sorted the data by charge state and re-plotted the graph with each charge state as a different series.

Observed mass vs delta error



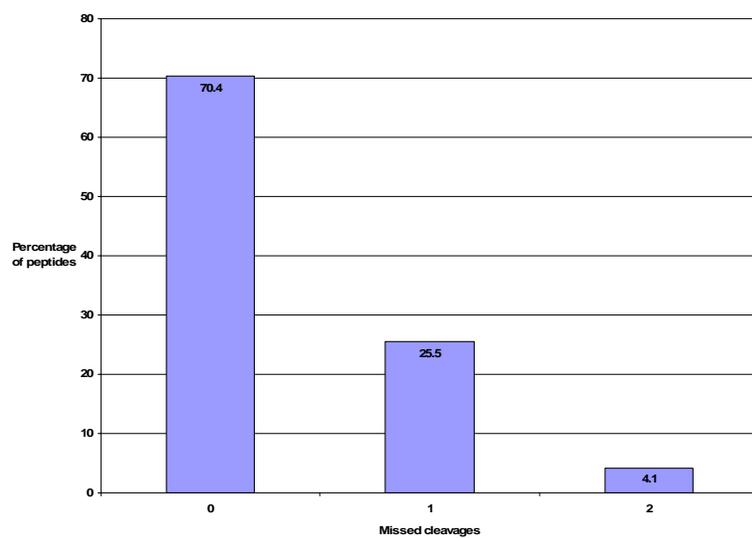
MASCOT : Problem solving with Mascot Integra

© 2006 Matrix Science

MATRIX
SCIENCE

And you can see that the three populations are indeed loosely related to charge state.

Missed cleavage distribution

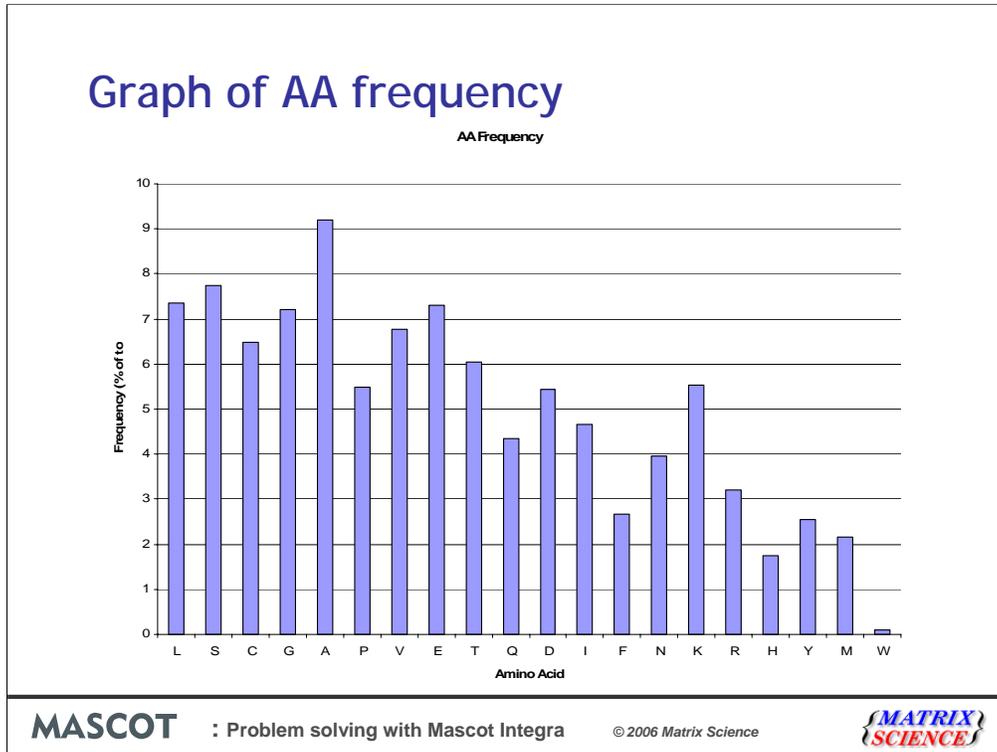


MASCOT : Problem solving with Mascot Integra

© 2006 Matrix Science



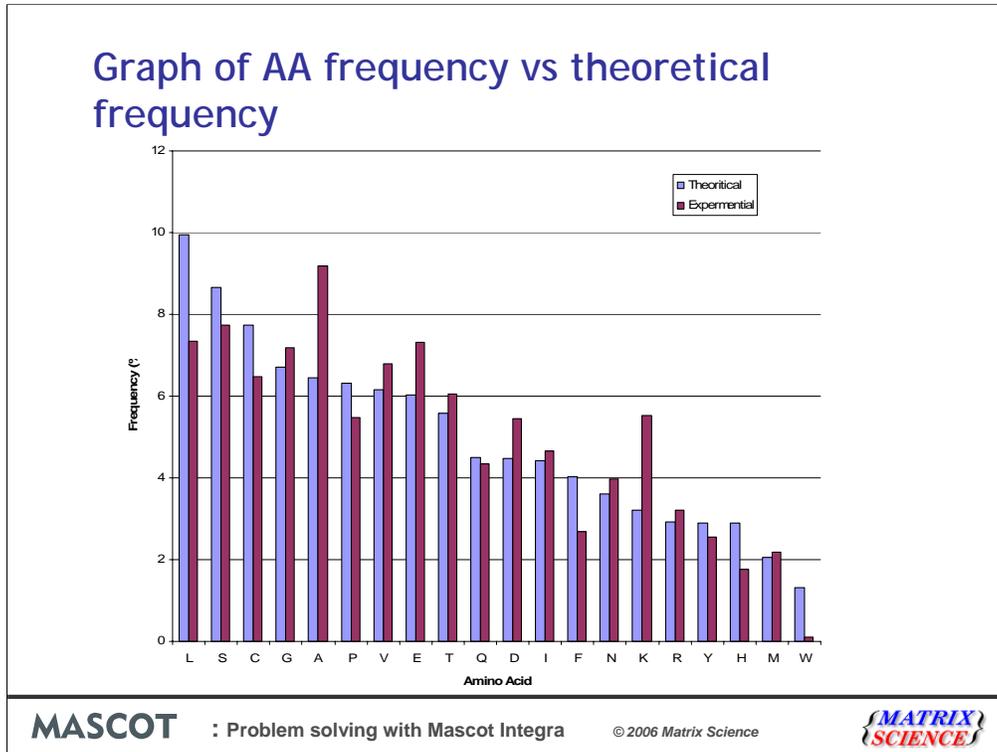
It was easy to build a missed cleavage histogram.



Next I calculated the amino acid frequency across all the peptides.

As this was an ICAT experiment the frequencies are going to be different compared to a complete proteome.

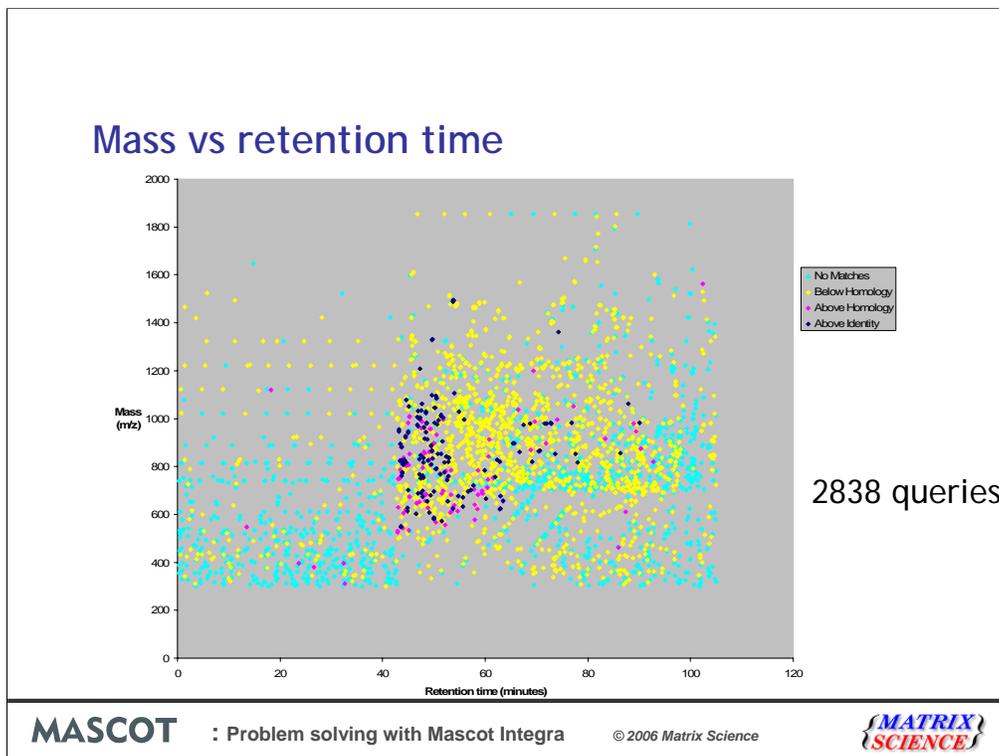
Therefore I generated the theoretical AA frequencies for all the tryptic Cysteine containing peptides and added them to the plot.



And we can see a number of differences between what was expected and what was measured.

Of particular interest was the under representation of Tryptophan.

I ran a number of error tolerant searches to see if I could identify more Tryptophan containing peptides with a consistent modification for example oxidation but the results have so far been inconclusive.



The second example is an overview plot of the mass and retention time of all the peptides analyzed by MS/MS from a single MuDPIT fraction.

There were 2838 ms/ms queries generated in the analysis of the fraction.

Four SQL queries have been combined on to one excel sheet to generate the four series shown in the plot.

The queries with a significant Mascot score are shown in blue.

The queries with a Mascot score greater than homology but less than the significant cut off are shown in purple

The queries with an insignificant match are shown in yellow

While the queries that did not generate a match are shown in cyan.

As you can see the majority of identifications eluted in the 40 to 80 minute period.

Protein vs DNA (human genome) data comparison

- A8 K562 erythroleukemia cells dataset processed using Mascot Distiller (38608 queries)
- Searched against IPI Human and Human genome databases
- Custom report designed to pull all significant peptide matches found in the Human genome search with no equivalent sequence match (above or below the threshold) in the IPI Human search
- Looking for novel peptide matches, SNPs, sequencing mistakes etc.

And the third example is a comparison between searching the IPI database and Human genome DNA database

We used the A8 Dataset from Katheryn Resing, and processed the raw data using Mascot Distiller to produce a dataset with 38608 queries. We then searched this against the IPI Human and Human Genome databases, and created a custom reports to identify all peptide matches from the Human genome search above the homology threshold with no equivalent sequence match (whether above the homology threshold or not) in the IPI Human search. By doing this we produce a list of peptides unique to the human genome search which we can use to look for novel peptide matches, possible sequencing errors, SNPs and other polymorphisms

1	QUERY TITLE	OBSERVED_MASS	OBSERVED_CHARGE	INTENSITY	MISSED_CLEAVES	MASCOT_SCORE	PEPTIDE_SEQUENCE	READABLE_VARS	MODS	BLAST
2	Sum of 2 scans in i	700.05703	2	19899276	0	69.58	ELEEIVQPLSK			http://www.ncbi.nlm.nih.gov/blast/
3	Scan 1339 (rt=550)	710.61938	2	202404	0	63.05	NAVEEYVYEMR	Oxidation (M)		http://www.ncbi.nlm.nih.gov/blast/
4	Sum of 2 scans in i	718.7216	2	2324639	1	49.49	IKWSDAGAEIVPK			http://www.ncbi.nlm.nih.gov/blast/
5	Sum of 3 scans in i	737.4399	2	6374620	0	86.32	QEAIIDLWQWR			http://www.ncbi.nlm.nih.gov/blast/
6	Scan 642 (rt=1927)	750.67315	2	9816316	0	74.43	WLHNEIQMAVEK			http://www.ncbi.nlm.nih.gov/blast/
7	Scan 846 (rt=2836)	798.26491	2	2429994	1	68.59	YVMDFEGEMKPKGR			http://www.ncbi.nlm.nih.gov/blast/
8	Sum of 2 scans in i	816.74939	2	1451900	1	71.18	QISEEPTNIMFAIR	Oxidation (M)		http://www.ncbi.nlm.nih.gov/blast/
9	Sum of 2 scans in i	840.81787	2	18411256	1	102.95	QIMKRGFLDAER			http://www.ncbi.nlm.nih.gov/blast/
10	Sum of 3 scans in rs	866.84857	2	8147765	0	105.41	AAPTAASDQPSAATTEK			http://www.ncbi.nlm.nih.gov/blast/
11	Scan 1186 (rt=6046)	447.09586	3	930407	1	73.73	NOMGIMEIKNK	2 Oxidation (M)		http://www.ncbi.nlm.nih.gov/blast/
12	Sum of 2 scans in i	923.69799	2	11646849	1	46.48	WPEVDDSPDLSEVK			http://www.ncbi.nlm.nih.gov/blast/
13	Sum of 5 scans in i	928.94056	2	9776327	0	89.35	NQDMMLSESNMPWFK			http://www.ncbi.nlm.nih.gov/blast/
14	Sum of 2 scans in i	958.49763	2	15129111	0	85.48	LVGSPFDPTTEGGPQVEK			http://www.ncbi.nlm.nih.gov/blast/
15	Scan 549 (rt=2724)	956.6758	2	1545095	1	88.85	LSEVEEADEASMTDPKPK			http://www.ncbi.nlm.nih.gov/blast/
16	Scan 1021 (rt=5168)	1009.357	2	6989138	0	86.35	TPMSEVLQAGGSMMDGPGPR			http://www.ncbi.nlm.nih.gov/blast/
17	Sum of 2 scans in i	1075.2316	2	7369796	0	76.72	GFQFYVYATVEVDVAVNNAK			http://www.ncbi.nlm.nih.gov/blast/
18	Sum of 6 scans in i	539.8557	2	43119027	0	61.41	PSANMPWFK			http://www.ncbi.nlm.nih.gov/blast/
19	Scan 959 (rt=2280.1)	1081.9632	2	2668635	2	101.89	EOSSEAAETGVSENEENPVR			http://www.ncbi.nlm.nih.gov/blast/
20	Sum of 2 scans in rs	1301.4832	1	7447788	0	69.85	MTTVHATATGK			http://www.ncbi.nlm.nih.gov/blast/
21	Sum of 2 scans in i	1369.9333	3	6696907	0	55.21	GLVPRGTEHEEEAPCCSSSSVGGAAASSPFAAGIPQEPGR			http://www.ncbi.nlm.nih.gov/blast/
22	Sum of 2 scans in i	1369.4412	2	715640	0	72.33	ATAPVPTVGEVGYGHESELSQASAAAR			http://www.ncbi.nlm.nih.gov/blast/
23	Sum of 2 scans in i	1492.2256	2	5472659	1	99.11	IVKAAENEYQTAISENYQTMSDTTFK			http://www.ncbi.nlm.nih.gov/blast/
24	Sum of 2 scans in i	1546.3083	1	2345814	0	102.24	AYHEQLSVAEINA			http://www.ncbi.nlm.nih.gov/blast/
25	Scan 952 (rt=5119)	1550.4317	2	5241041	1	73.63	YTLPPGVDPYVSSLSPEGLTVEAPMPK			http://www.ncbi.nlm.nih.gov/blast/
26	Sum of 2 scans in i	1614.4634	2	4080166	2	86.56	KYTLPPGVDPYVSSLSPEGLTVEAPMPK			http://www.ncbi.nlm.nih.gov/blast/

- 25 distinct peptide sequences above the homology or significance threshold identified in the Human Genome search with no equivalent match from the IPI Human search
- Take a closer look at one of these sequences:
 - IVKAAENEYQTAISENYQTMSDTTFK

Running the report on these samples identified 25 distinct peptide sequences which were above the homology threshold in the Human Genome search which were absent from the IPI Human search. Doing the comparison by hand would have taken many hours (or days), but by having all of the information stored inside Mascot Integra, we were able to generate this hit list in a couple minutes.

The next step is, of course, to see if any of these peptides are telling us anything interesting – do any of them represent novel sequences, polymorphisms or sequencing errors? We'll take a closer look at one of the peptides identified.

IVKAAENEYQTAISENYQTMSDITTFK

Top BLAST match (against NCBI nr) :

Have a K->E in the nr sequence compared with the Human genome match

IVEAAENEYQTAISENYQTMSDITTFK was found in IPI Human search as a match to the same query

```
>gi|62897569|db|BAD96724.1| capping protein (actin filament) muscle 2-line, alpha 2 variant [Homo sapiens]
gi|62897539|db|BAD96709.1| capping protein (actin filament) muscle 2-line, alpha 2 variant [Homo sapiens]
Length=286

Score = 82.9 bits (188), Expect = 1e-15
Identities = 25/26 (96%), Positives = 26/26 (100%), Gaps = 0/26 (0%)

Query 1  IVKAAENEYQTAISENYQTMSDITTFK 26
          IVKAAENEYQTAISENYQTMSDITTFK
Sbjct 231 IVEAAENEYQTAISENYQTMSDITTFK 256
```

Protein hit list

ID	Accession	Description	MS Rank	Mascot Protein Score
<input type="checkbox"/>	msh-2042096	ms-2042096_P-actin capping protein	400	195
<input type="checkbox"/>	P00011	alpha-2 subunit	400	195

MASCOT : Problem solving with Mascot Integra

© 2006 Matrix Science



The first step was to BLAST the peptide sequence against NCBI nr and see if there is a match in the database. In this case we retrieved a good match to an actin capping protein – the only difference between the sequence from the Human Genome search against the sequence in NCBI nr is the K->E at position 3 in the NCBI nr sequence compared with the Human genome match.

Since the published sequence of this peptide/protein has Glutamic Acid at position 3, then any match from the IPI Human search would have been to this sequence. If we do a search in Integra for any protein hits containing this sequence, then we do indeed find a match to the peptide from the same query in the IPI Human search.

Match to Human genome:

Ions Score: 99 Expect: 0.00004344
Matches (Bold Red): 16/282 Fragment ions using 16 most intense peaks

#	b	b ⁺	b ⁺	b ⁺⁺	b ⁺	b ⁺⁺	Seq.	y	y ⁺	y ⁺	y ⁺⁺	y ⁺	y ⁺⁺	#
1	114.09	57.55					I							26
2	213.16	107.08					V	2869.32	1435.16	2852.29	1426.65	2851.31	1426.16	25
3	341.25	171.13	324.23	162.42			K	2770.25	1385.63	2753.22	1377.12	2752.24	1376.62	24
4	412.29	206.45	395.27	196.14			A	2642.15	1321.59	2625.13	1313.07	2624.15	1312.55	23
5	483.23	242.17	466.3	233.45			A	2571.12	1286.06	2554.09	1277.55	2553.11	1277.04	22
6	612.37	306.69	595.34	298.18	594.36	297.68	E	2500.08	1250.54	2483.06	1242.03	2482.07	1241.54	21
7	726.41	363.71	709.39	352.2	708.4	354.71	N	2371.04	1186.02	2354.01	1177.51	2353.03	1177.02	20
8	855.46	428.23	838.43	419.22	837.45	419.23	E	2257	1129	2239.97	1120.49	2238.99	1120	19
9	1018.82	509.74	1001.49	501.25	1000.51	500.74	Y	2127.95	1064.48	2110.93	1055.97	2109.94	1055.49	18
10	1146.58	573.79	1129.55	565.28	1128.57	564.79	Q	1964.89	982.95	1947.86	974.44	1946.88	973.94	17
11	1247.63	624.32	1230.6	615.8	1229.62	615.31	T	1836.83	918.92	1819.81	910.41	1818.82	909.91	16
12	1318.66	659.84	1301.64	651.32	1300.65	650.83	A	1735.78	868.4	1718.76	859.88	1717.77	859.29	15
13	1451.75	725.88	1434.72	727.36	1433.74	727.37	I	1664.75	832.88	1647.72	824.36	1646.74	823.87	14
14	1518.78	759.89	1501.75	751.38	1500.77	750.89	S	1551.66	776.34	1534.64	767.82	1533.65	767.33	13
15	1647.82	824.41	1630.8	815.9	1629.81	815.41	E	1464.63	732.82	1447.6	724.31	1446.62	723.81	12
16	1761.87	881.44	1744.84	872.92	1743.85	872.43	N	1335.59	668.3	1318.56	659.78	1317.58	659.29	11
17	1924.93	962.47	1907.9	954.45	1906.92	953.96	F	1221.53	611.28	1204.52	602.76	1203.54	602.27	10
18	2052.99	1027	2035.96	1018.48	2034.98	1017.99	Q	1058.48	529.24	1041.46	521.23	1040.47	520.74	9
19	2154.98	1077.52	2137.01	1069.01	2136.02	1068.52	T	930.42	465.72	913.4	457.2	912.41	456.71	8
20	2285.08	1143.04	2268.05	1134.53	2267.06	1134.04	M	829.38	415.19	812.35	406.68	811.37	406.19	7
21	2372.11	1186.56	2355.08	1178.04	2354.1	1177.95	S	690.34	349.67	681.31	341.16	680.32	340.67	6
22	2467.13	1234.07	2450.11	1225.56	2449.12	1225.07	D	611.3	306.16	594.28	297.64	593.29	297.15	5
23	2588.18	1294.59	2571.16	1286.08	2570.17	1285.59	T	496.28	248.64	479.25	240.13	478.27	239.64	4
24	2689.23	1345.12	2672.2	1336.61	2671.22	1336.11	T	395.23	198.12	378.2	189.6	377.22	189.11	3
25	2836.2	1418.65	2819.27	1410.14	2818.29	1409.65	F	294.18	147.59	277.15	139.08			2

Match to IPI Human:

Ions Score: 123 Expect: 0
Matches (Bold Red): 31/280 Fragment ions using 37 most intense peaks

#	b	b ⁺	b ⁺	b ⁺⁺	b ⁺	b ⁺⁺	Seq.	y	y ⁺	y ⁺	y ⁺⁺	y ⁺	y ⁺⁺	#
1	114.09	57.55					I							26
2	213.16	107.08					V	2870.27	1435.64	2853.24	1427.12	2852.26	1426.63	25
3	342.2	171.6					E	2791.2	1396.1	2754.17	1377.59	2753.19	1377.1	24
4	412.29	207.12					A	2642.14	1321.59	2625.13	1313.07	2624.15	1312.55	23
5	484.28	242.64					A	2571.12	1286.06	2554.09	1277.55	2553.11	1277.04	22
6	612.32	307.16					E	2500.08	1250.54	2483.06	1242.03	2482.07	1241.54	21
7	727.36	364.18	710.34	356.67	709.35	355.18	N	2371.04	1186.02	2354.01	1177.51	2353.03	1177.02	20
8	856.4	428.21	839.38	420.19	838.39	419.7	E	2257	1129	2239.97	1120.49	2238.99	1120	19
9	1019.47	510.24	1002.44	501.72	1001.46	501.23	Y	2127.95	1064.48	2110.93	1055.97	2109.94	1055.49	18
10	1147.53	574.27	1130.5	565.75	1129.52	565.26	Q	1964.89	982.95	1947.86	974.44	1946.88	973.94	17
11	1248.57	624.79	1231.55	616.28	1230.56	615.79	T	1836.83	918.92	1819.81	910.41	1818.82	909.91	16
12	1318.61	660.31	1301.58	651.8	1300.6	651.3	A	1735.78	868.4	1718.76	859.88	1717.77	859.29	15
13	1432.7	716.85	1415.67	708.34	1414.68	707.85	I	1664.75	832.88	1647.72	824.36	1646.74	823.87	14
14	1519.73	760.37	1502.7	751.85	1501.72	751.36	S	1551.66	776.34	1534.64	767.82	1533.65	767.33	13
15	1648.77	824.89	1631.74	816.38	1630.76	815.88	E	1464.63	732.82	1447.6	724.31	1446.62	723.81	12
16	1762.81	881.91	1745.79	873.4	1744.8	872.9	N	1335.59	668.3	1318.56	659.78	1317.58	659.29	11
17	1925.88	963.44	1908.85	954.93	1907.87	954.44	Y	1221.53	611.28	1204.52	602.76	1203.54	602.27	10
18	2053.93	1027.47	2036.91	1018.96	2035.92	1018.47	Q	1058.48	529.24	1041.46	521.23	1040.47	520.74	9
19	2154.98	1077.99	2137.96	1069.48	2136.97	1068.99	T	930.42	465.72	913.4	457.2	912.41	456.71	8
20	2286.02	1143.52	2269	1135	2268.01	1134.51	M	829.38	415.19	812.35	406.68	811.37	406.19	7
21	2373.05	1187.03	2356.03	1178.52	2355.04	1178.03	S	690.34	349.67	681.31	341.16	680.32	340.67	6
22	2488.08	1244.54	2471.06	1236.03	2470.07	1235.54	D	611.3	306.16	594.28	297.64	593.29	297.15	5
23	2589.13	1295.07	2572.1	1286.56	2571.12	1286.06	T	496.28	248.64	479.25	240.13	478.27	239.64	4
24	2690.18	1345.59	2673.15	1337.08	2672.17	1336.59	T	395.23	198.12	378.2	189.6	377.22	189.11	3
25	2837.25	1419.13	2820.22	1410.61	2819.24	1410.12	F	294.18	147.59	277.15	139.08			2

- Match to IPI Human more likely to be the correct match to the query
- Sequence in Human genome database either incorrect or represents a polymorphism, or we have a spurious match (which could be a match to a homologue)

If we take a closer look at the peptide matches...Here we have the match from the Human Genome search – its a pretty good match with a reasonable run of y ions and gets a good score (98). However, the match to IPI Human (with Glutamic Acid at position 3) is clearly the stronger match – we have additional b series matches and a higher ions score.

Therefore, we would conclude that the match to the sequence published in IPI Human and NCBI is probably the correct match to this query. The sequence in the Human Genome database is either incorrect (i.e. a sequencing error) or it represents a polymorphism, or we have a spurious match. The spurious match could be a genuine ‘false positive’, or it could be a match to an homologous protein sequence. If it is to a homologous sequence then the analysis could have identified a match of interest.

Protein vs DNA (human genome) data comparison

Running the report the other way round we find 1083 peptides with scores above the homology threshold specific to IPI Human

Represents a loss of 24.2% of peptides in the Human genome search. Close to what we would expect to lose at Intron/Exon boundaries.

MASCOT

: Problem solving with Mascot Integra

© 2006 Matrix Science

MATRIX
SCIENCE

The report can, of course, be run the other way around, to identify peptide matches unique to the IPI Human search. If we do this we find 1766 peptides from the IPI Human search with scores above the homology threshold which do not have an equivalent match in the Human genome search.

The IPI Human search identified a total of 4461 possible peptide matches with scores above the homology threshold, so the 1083 'missing' peptides represent approximately 24% of the total number of peptides identified from IPI Human – close to the approx 20% of peptide matches we would expect to lose because of Intron/Exon boundaries.

Advanced reporting

96 well plate assay report.

PMF report that satisfies MCP guidelines.

- Blast cluster of proteins to determine nearest non homologous match
- Meet all of the MCP criteria

Now on to more advanced reporting. As I said earlier we could use many different computer languages to do this but for these examples I used the perl scripting language.

The first example is of an assay report for a 96 well plate.

The second is a report for the proteins identified from a 2D gel by peptide mass fingerprinting and it has to satisfy the Molecular and Cellular proteomics guide lines

96 well plate assay report

The screenshot displays the Mascot software interface. On the left, the 'plateReporter' window shows a list of 96 well plates, with 'P-060300210' selected. The main window shows an Excel spreadsheet with columns for 'Well', 'MS File Name', 'MS Database ID used by Search Engine for seq', 'Score', 'Peptide Seq.', 'Coverage', 'Sequence', 'Coverage', and 'Link'. The table lists various protein identifications such as 'Tryptic Digest', 'MS0000010', 'MS0000040', etc. At the bottom, the 'MASCOT' logo is visible along with the text 'Problem solving with Mascot Integra © 2006 Matrix Science' and the 'MATRIX SCIENCE' logo.

Imagine that you have an assay with a single protein per a well of 96 well plate and that you are interested in the modified peptides for each protein.

First we select the plate.

Then advanced reporting application pulls out the top protein result for each well of the plate and list the modified peptides for the protein with colour coded amino acids and save the results into an excel sheet.

Molecular Cellular Proteomics publication guidelines

Guideline 1: Supporting information

- 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.
- The name and version of the sequence database(s) used.

This next example generates a peptide mass fingerprint report that meets the Molecular Cellular Proteomics publication guidelines.

I'm only going to review the guidelines that pertain to general reporting or are specifically for PMF data.

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

The data analysis program and parameters used

The database searching program and parameters used

And the name and version of the database used.

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.

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 second guideline specifies the protein information that should be reported while the third guideline is concerned with the peptides.

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

Guideline 6 is for peptide mass fingerprinting and asks for 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.

MCP PMF report guideline 1

	A	B	C
1	EXP-060400266		
2	Peak picking parameters		
3	Peak picking program	MDRO (Mascot Distiller) 2.0.0.0	
4	MS.AggregationMethod	1	
5	MS.MaxPeakCharge	1	
6	MS.MinPeakCount	1	
7	MS.RegriddingPointsPerDa	50	
8	MS.UncentroidingHalfWidth	0.08	
9	MS.UncentroidingPointsPerDa	50	
10	MSMS.AggregationMethod	2	
11	etc . . .		
12			
13	Database Search Parameters		
14	Search engine	Mascot	
15	Database	Sprot	
16	Database Size	212425	
17	Taxonomy Fungi	
18	Database Size after Taxonomy	14157	
19	Mass Accuracy	100	
20	Missed cleavages	2	
21	Modifications	Carbamidomethyl (C)	
22	Enzyme	Trypsin	
23	Resolution	N/A	
24	Calibration	N/A	
25	Exclusion	N/A	
26			

MASCOT : Problem solving with Mascot Integra

© 2006 Matrix Science



After selecting the 2D gel experiment the application extracts the processing and database searching parameters from Mascot Integra.

MCP PMF report guidelines 2 and 6 - proteins

SearchID	Protein Accession No	Protein Description	Mascot Score	Mascot Probability	Next best hit Mascot Score	Next best hit Mascot Probability	Number matched peaks	Number unmatched peaks	Sequence Coverage	Number unique peptides	Protein Mass pI	
52	mss-25042006-00046	HSP70_SCHPO (P04621) Heat shock protein 70 kDa	49387	4.93875E4	463	0.331872475	15	79	77.16	12	17762.42 9.41	
53	mss-25042006-00046	MDM1_SCHPO (Q43003) Protein mdm1, intracell	94.6	4.93875E4	NA	NA	7	41	16.11	7	43411.83 5.3	
54	mss-25042006-00047	GPD1_SACBA (Q6JL53) Glycerol-3-phosphate 1-phosphatidyltransferase	38.8	1.86629E08	NA	NA	7	103	11.4	5	23065.39 6.23	
55	mss-25042006-00048	RHOA_EMENI (Q9CY44) Protein rhoA (Rho) pt	21.1	109.89330	NA	NA	5	81	16.46	10	39197.77 9.79	
56	mss-25042006-00049	YC19_YEAST (P53328) Hypothetical 37.9 kDa	39.4	2.0463300E	NA	NA	10	19	12.04	7	34854.8 6.46	
57	mss-25042006-00050	TH4_FUSH (P23617) Thiazole biosynthetic t	62.9	0.0072656	49.3	0.186330247	7	87	9.27	5	17244.46 5.36	
58	mss-25042006-00051	UBC2_YARLI (Q6C953) Ubiquitin-conjugating e	27.9	22.9599669	NA	NA	5	69	45.33	42	71752.71 5.29	
59	mss-25042006-00052	PABP_SCHPO (P31209) Poly(ADP-ribose) binding	221		43.9	0.576720259	49	6	113	17.89	5	16500.44 6.89
60	mss-25042006-00053	UBCA_CANAL (P43102) Ubiquitin-conjugating e	23.9	57.6720259	NA	NA	6	99	21.21	10	34619.73 8.95	
61	mss-25042006-00054	RRP7_YEAST (P25368) Ribosomal RNA-proce	28.6	19.542100E	NA	NA	10	59	12.04	7	34854.8 6.46	
62	mss-25042006-00055	TH4_FUSH (P23617) Thiazole biosynthetic t	36.2	3.3952077E	NA	NA	7	119	62.96	7	11827.86 11.7	
63	mss-25042006-00056	RL3C_TRHM (Q8HFR7) GDS ribosomal proten	35.3	4.1782060E	NA	NA	7	54	23.71	6	20322.24 11.09	
64	mss-25042006-00057	GARI_SCHPO (Q66975) HAAC ribonucleoproti	36.7	3.026712E	NA	NA	6	79	33.51	8	20322.24 11.09	
65	mss-25042006-00058	GARI_SCHPO (Q66975) HAAC ribonucleoproti	42	0.8924651E	NA	NA	8	93	15.38	11	59502.79 5.71	
66	mss-25042006-00059	YME3_YEAST (Q04712) Hypothetical 59.3 kDa	36.4	3.2431813E	NA	NA	11	48	17.59	8	24798.79 9.34	
67	mss-25042006-00060	CWC24_CANAL (Q6ACW2) Pre-mRNA-splicing f	32.1	8.7291264E	NA	NA	8	70	19.5	6	32254.41 6.63	
68	mss-25042006-00061	PSOR_SCHPO (Q89717) Pyruvate-C-carboxylat	26.9	29.576185E	NA	NA	6	50	12.65	7	34854.8 6.46	
69	mss-25042006-00062	TH4_FUSH (P23617) Thiazole biosynthetic t	39.5	1.5884415E	NA	NA	7	63	13.82	3	18861.2 4.66	
70	mss-25042006-00063	DAP1_YEAST (O12091) Damage response pro	16.5	316.93575E	NA	NA	3	43	59.96	30	67449.01 5.82	
71	mss-25042006-00064	HSP70_SCHPO (O10286) Heat shock protein sk	220		47.8	0.224947718	34	48	75.9	36	64596.13 6.36	
72	mss-25042006-00065	PURS_SCHPO (O74928) Bifunctional purine bio	271		50.7	0.120495612	44	75	62.2	34	55244.45 4.5	
73	mss-25042006-00066	PDH_SCHPO (O10057) Putative protein desulf	238		56	0.035660776	41	34	0	31	0	
74	mss-25042006-00067	Mixture 1	303		42.6	0.777969015	50	52	57.51	31	69218.75 5.8	
75	mss-25042006-00068	VATA_SCHPO (P31406) Vacuolar ATP synthas	217		42.6	0.777969015	32	66	41.92	17	67449.01 5.82	
76	mss-25042006-00069	HSP70_SCHPO (O10286) Heat shock protein sk	77.3	0.00026361	42.6	0.777969015	18	52	39.63	26	60717.18 4.89	
77	mss-25042006-00070	HSP90_SCHPO (P41887) Heat shock protein 90	106	3.55038E4	38.4	2.0463300E3	30	50	32.01	23	60717.18 4.89	
78	mss-25042006-00071	HSP90_SCHPO (P41887) Heat shock protein 90	65	4.47684E4	40.9	1.150724162	24	83	71.78	53	67449.01 5.82	
79	mss-25042006-00072	HSP70_SCHPO (O10286) Heat shock protein sk	369		46.5	0.316935752	59	121	84.6	38	67449.01 5.82	
80	mss-25042006-00073	HSP70_SCHPO (O10286) Heat shock protein sk	167	2.825E-1	58.6	0.0195421	40	44	39.8	14	67449.01 5.82	
81	mss-25042006-00074	HSP70_SCHPO (O10286) Heat shock protein sk	74	0.00056E	34.1	5.507712118	14	59	70.96	46	62413.88 5.76	
82	mss-25042006-00075	HSP60_SCHPO (Q02864) Heat shock protein 60	330		55.9	0.036389093	49					

MASCOT : Problem solving with Mascot Integra

© 2006 Matrix Science



Then reports the top ranked protein hit along with information required by guidelines 2 and 6 to the first worksheet.

MCP PMF report guidelines 3 - peptides

QueryID	Observed m/z	M(calc)	M(calc) Delta	Intensity	Start residue	End residue	Missed cleavages	Peptide sequence	Variable modifications
217	717.3684	716.361124	716.41806	19882.226	2	7	0	1 M.TNINLKK.T	
219	725.26811	724.250834	724.296234	137373.66	195	199	0	0 M.NCEFR.I	
220	764.42909	763.421814	763.459198	190479.38	8	13	0	0 K.TVYLIR.H	
222	864.45435	863.447074	863.453461	5777.1526	1	7	1	1 -.MTNINLKK.T	Oxidation (M)
223	892.5418	891.534524	891.554153	4830.1898	7	13	1	1 K.KTYLIR.H	
224	949.47755	948.470274	948.455215	12064.02	78	86	0	0 K.YLAEGGPK.V	
225	1147.6462	1146.638924	1146.631821	4281.8448	67	76	0	0 R.TLOTMEIALK.K	
226	1163.6471	1162.639824	1162.62674	3265.5197	67	76	0	0 R.TLOTMEIALK.K	Oxidation (M)
227	1275.7498	1274.742524	1274.726776	16357.291	67	77	1	1 R.TLOTMEIALK.K.Y	
228	1291.7563	1290.749024	1290.721695	13297.238	57	77	1	1 R.TLOTMEIALK.K.Y	Oxidation (M)
229	1423.7984	1422.791124	1422.735426	186174.34	134	146	0	0 R.DVIASDVTSAIR.S	
230	1462.8104	1461.803124	1461.866699	3506.2947	2	13	2	2 M.TNINLKK.TVYLIR.H	
231	1480.8283	1479.821024	1479.782043	227614.64	200	212	0	0 R.IYDLVGTITSELK.L	
232	1485.8001	1484.792824	1484.747925	49797.273	53	65	0	0 K.QIPIDGVCSPMR.R	
233	1501.7915	1500.784224	1500.742844	39930.012	53	65	0	0 K.QIPIDGVCSPMR.R	Oxidation (M)
234	1561.8667	1560.859424	1560.818756	484212.39	181	194	0	0 K.AADIDFPPQLSFK.N	
235	1641.8938	1640.886524	1640.84903	41334.572	53	66	1	1 K.QIPIDGVCSPMRR.T	
236	1657.8804	1656.863124	1656.843948	62731.983	53	66	1	1 K.QIPIDGVCSPMRR.T	Oxidation (M)

MASCOT : Problem solving with Mascot Integra

© 2006 Matrix Science



On a second worksheet the peptide information for each protein hit is reported and that satisfies guideline 3.

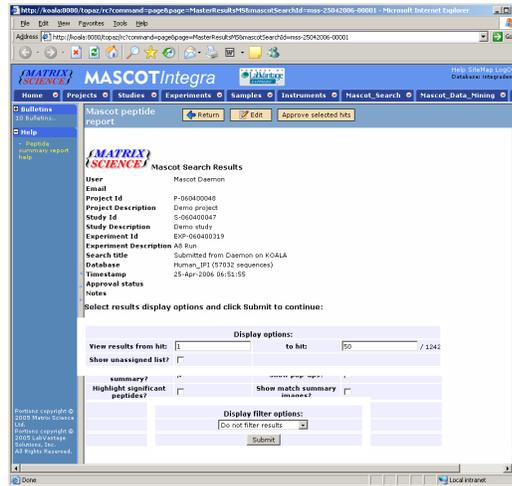
4. Large search viewing

Paging of results

- Can select range of protein hit ranks to view
- Far easier to handle than all 1242 protein hit ranks at once
- Avoids client side browser memory issues
- Much faster to generate report
 - 5 hit ranks ~ 9seconds
 - 50 hit ranks ~ 20 seconds
 - 100 hit ranks ~ 30 seconds
 - All 1242 ranks ~ 11 minutes*

Filtering of results

- Integra ships with a range of filters which can be applied to reports
- Can write your own SQL filter to limit proteins returned
- Can filter on any aspect of the protein, peptide or original query.



* when opened either from Mascot Integra or from Mascot server

MASCOT : Problem solving with Mascot Integra

© 2006 Matrix Science



Another advantage of using Mascot Integra is that we can simplify viewing large result sets in the standard reports.

For example because we are pulling the protein hits out of the Integra database we can specify a range of protein hit ranks to view in the report. For example here I'm only going to open up the first 50 protein hit rank from the A8 Dataset – a report which contains 1242 protein hit ranks in total. This means that we only have to deal with a subset of the results at any one time, and that Internet Explorer isn't going to run out of memory on your client PC – which can be a major problem when opening large result sets – and the report is much faster to generate than opening all of the results on the Mascot server. For example, opening just 5 hit ranks from the A8 dataset takes 9 seconds, 20 seconds to open 50, 30 seconds to open 100 but approximately 11 minutes to open all 1242 hit ranks. Opening the report from the Mascot server also took approximately 11 minutes (the results started displayed earlier, but it was 11 minutes before IE was accepting user input)

We can also write SQL filters which produce a valid list of protein hits to display. We can filter on any aspect of the protein hit

- i.e. the protein length, mass or pl – on the peptides which the protein has matched
- i.e. require that the hit contains at least X distinct peptide matches above the homology threshold – or on the original query – i.e. pull out all the proteins which contain a good match to Query 53

The screenshot shows the Mascot Integra web interface in a Microsoft Internet Explorer browser. The page title is "Mascot peptide report". The main content area displays "Mascot Search Results" with a summary of the search parameters:

- User: Mascot Daemon
- Email: Mascot Daemon
- Project Id: P-060400048
- Project Description: Demo project
- Study Id: S-060400047
- Study Description: Demo study
- Experiment Id: EXP-060400319
- Experiment Description: A6 Run
- Search title: Submitted from Daemon on KOALA
- Database: Human_IP1 (57032 sequences)
- Timestamp: 25-Apr-2006 06:51:55
- Approval status: Notes

Below the search results, there are two red circles highlighting specific elements:

- The first circle highlights the "Display options:" section, which includes a "View results from hit:" field set to "5" and a "to hit:" field set to "1242 / 1242".
- The second circle highlights the "Display filter options:" dropdown menu, which is currently set to "Do not filter results".

To the right of the main interface, a separate window titled "Display filter options:" is shown, listing various filter criteria:

- Do not filter results
- C-term matches
- N-term matches
- Minimum Mass X
- Minimum Coverage X
- Maximum mass X
- Pep. must match XXXXX
- X sig at Y threshold
- X sig with range
- Minimum pI
- Maximum pI
- By Accession

At the bottom of the page, there is a footer with the Mascot logo and the text: "MASCOT : Problem solving with Mascot Integra © 2006 Matrix Science".

This is how we apply a filter

Note that I've chosen to display all 1242 hit ranks. This is because the result filter and view to/from ranks work in conjunction. If I had chosen to display only the 1st 50 protein hit ranks, then the filter would only work on those ranks.

1st we select our filter from the drop down list. Here we're using one of the filters that comes with Mascot Integra, which allows us to filter out any proteins which do not contain a Mascot peptide match to a peptide sequence/subsequence.

The screenshot shows the Mascot search interface. A dialog box titled "Explorer User Prompt" is open, asking for a "Script Prompt" with the text "peptide_sequence like" and a red circle around the input field containing "QEAIQDLVQWR". The main window displays search results for "peptide_sequence like QEAIQDLVQWR" with 1242 hits. A "Peptide Select summary report" is shown for the top hit, Nucleophosmin (Ta1_192032240). The report lists various peptides and their scores, with hit number 7 being the only one that matches the filter. The search parameters are also displayed at the bottom.

Search Parameters

Type of search:	MS/MS Ion Search
Enzyme:	Trypsin/P
Fixed Modifications:	Carbamidomethyl (C)
Variable Modifications:	Oxidation (O)
Mass values:	Monoisotopic
Database:	Human (NCBI)

When we select the filter, we are presented with a popup window into which we must enter the peptide subsequence we are interested in. In this case we'll use one of the peptide sequences identified in the A8 dataset.

Click OK and then Submit.

This is the resulting report. Out of all 1242 protein hit ranks, only hit number 7 – Nucleophosmin – contains a match to the required peptide subsequence.

MASCOTIntegra : Problem solving with Mascot Integra © 2006 Matrix Science

Results filtered using Minimum pI (pI >= 9)

Peptide Select Summary report
 36: **1200211462** Mass: 21221 Total Score: 843 Peptides Matched: 50
 Tax_ID=9606 Histone H1.2
 Query Observed Mr(exp) Mr(Calc) Delta Miss Score Rank Peptide

Query Observed Mr(exp)	Mr(Calc)	Delta	Miss	Score	Rank	Peptide		
126	604.017	603.01	600.396	2.614	2	9	10	K.KTPEK.A
184	657.997	656.99	657.381	-0.391	1	5	5	K.AGTEPEK.K
5287	374.005	745.996	745.433	-0.562	0	25	1	K.GTIVQTK.G 5000 5004
2423	672.466	671.459	671.36	0.098	0	25	1	K.AAGGATPEK 2534 2561 2562
8228	818.044	817.036	813.482	3.554	2	13	3	K.KAGGTPK.A
7288	400.178	798.342	797.512	0.83	2	10	9	K.KPVGAEEK.P
2502	726.438	727.431	725.433	-0.003	0	40	1	K.PAAATPEK 2503
7387	407.059	812.104	810.387	1.717	0	35	1	K.GTGAAGSPK.L 2251
11734	888.573	887.565	885.528	2.037	1	1	2	K.KPAAATVTK.K
22210	541.995	1083.974	1084.66	-2.666	2	7	4	K.AKPPAATVTK
25319	554.877	1107.74	1106.561	1.179	0	78	1	K.ALAAGYVEK.N 25251 25262 25263 25271 25287 25295 25308 25347
29437	600.281	1198.548	1197.66	0.887	0	58	1	K.ASGPPVSEITK.A 3 23391 23417
30433	1235.566	1234.559	1234.656	-0.097	1	71	1	K.KALAAAGYVEK.N
4201	663.926	1325.837	1325.725	0.082	1	116	1	K.ASGPPVSEITK.A 2205 2212 11572 11614 11615 11628 11629 11646 33023 33025
4806	740.211	1478.407	1477.741	0.666	0	99	1	-SETAPAAPAAPPEK.A 4732 4732 12738 36085 36088
8324	830.39	1658.766	1656.952	1.814	2	6	6	K.AVAASKERSGVLAALK.K
19265	1006.78	2011.544	2010.111	1.434	2	3	3	K.ASGPPVSEITKVAASKER.S

41: **1200211467** Mass: 21720 Total Score: 821 Peptides Matched: 51
 Tax_ID=9606 Histone H1.4
 Query Observed Mr(exp) Mr(Calc) Delta Miss Score Rank Peptide

Query Observed Mr(exp)	Mr(Calc)	Delta	Miss	Score	Rank	Peptide		
126	604.017	603.01	600.396	2.614	2	9	10	K.KTPEK.A
5287	374.005	745.996	745.433	-0.562	0	25	1	K.GTIVQTK.G 5000 5004
4004	716.486	715.479	715.386	0.093	0	24	1	K.ATGAATPEK 4000 4005
8235	415.379	828.744	825.507	3.237	2	9	8	K.KAPEPAE.A
9237	838.26	837.253	840.519	-3.265	2	3	2	K.AKPPGAEEK.K
10253	429.22	856.426	856.303	-2.072	2	2	10	K.AKPSGAEEK.R
7387	407.059	812.104	810.387	1.717	0	35	1	K.GTGAAGSPK.L 2251
9551	423.192	844.369	843.481	0.887	1	15	4	K.KATGAATPEK 3549 9551
10787	468.355	937.710	936.534	0.714	2	9	7	K.AKAPPEV.V

Some more quick examples of the type of things we can do with filtering.

1: Here we are filtering for pI and only viewing protein hits with a predicted pI of > 9

The screenshot shows the Mascot Integra web interface. The browser address bar displays the URL: `http://koad1000/topaz/rc/command-page?page=MasterResultM5`. The page title is "MASCOT Integra". The navigation menu includes: Home, Projects, Studies, Experiments, Samples, Instruments, Mascot_Search, Mascot_Data_Mining, and Utilities. The main content area shows search results for "Tax_ID=9686 Prefoldin subunit 5". The results are displayed in a table with columns: Query, Observed, Mr(calc), Mr(calc), Delta, Miss, Score, Rank, and Peptide. Several rows are highlighted in yellow, indicating significant matches. The table shows the following data:

Query	Observed	Mr(calc)	Mr(calc)	Delta	Miss	Score	Rank	Peptide
598	620.568	629.56	633.297	-3.737	0	16	2	K.TAEDAK.D
10284	975.67	974.663	974.485	0.177	0	4	5	K.DCLNVLK.S
31380	1271.592	1270.384	1267.648	2.936	1	3	3	K.IDFLTKQEK.I
7812	408.67	1198.989	1197.519	1.47	0	12	7	K.GAVSIEHNSK.I
2822	706.029	1416.043	1408.75	1.293	1	12	8	K.IQALQEKHAK.Q
9810	849.802	1697.59	1696.74	0.85	1	3	5	K.HANKQAVDENMSQK.I 9700
0147	814.733	1631.452	1632.08	-1.427	0	3	9	K.ELLWPLTSHYVPGK.L
20445	1857.681	2063.152	2062.088	1.194	0	157	1	K.NLDIEQVDFEELSTAKL.K 20415 20423 20474 20440
4874	735.322	2202.943	2201.144	1.8	0	6	1	-NAQSINTEINLPQLENL.K.N
3258	1348.806	2695.756	2695.334	0.422	1	0	9	K.QAVSIEHNSQK.IQTLGAQAATAK.A
21442	1838.58	3112.718	3114.581	-1.863	2	1	8	K.HANKQAVDENMSQK.IQTLGAQAATAK.A
22523	1606.114	4815.321	4816.524	-1.203	2	2	10	-NAQSINTEINLPQLENLKLQDQVEFLSTIAQLKVVQTE.Y

The interface also includes a "Display filter options" section with a dropdown menu set to "One hit wonders" and a "Submit" button. The footer of the page contains the Mascot logo, the text "Problem solving with Mascot Integra", the copyright notice "© 2006 Matrix Science", and the Matrix Science logo.

2: Here we filtering for only those proteins that contain one significant peptide match so called one hit wonders.

Display filter options:
With Modifications

Selection filter options:
Include keywords:
Exclude keywords:
Expectation value threshold: 0.05

Peptide summary report
1 Selected hit: gi|605340 Kaa123p: Karyopherin beta 4 [Saccharomyces cerevisiae]

Comments:
@ 605340 Mass: 12254 Total Score: 283 Peptides Matched: 12
Kaa123p: Karyopherin beta 4 [Saccharomyces cerevisiae]

Match#	Query	Observed	M(exp)	M(Calc)	Delta	Miss	Score	Rank	Peptide
2		357.7299	713.4453	713.4687	-0.0234	0	33	6	K.VIELLK.Y
3		359.2217	716.4288	716.4432	-0.0144	0	30	8	K.VLLASK.V
76		480.7859	959.5573	959.5691	-0.0118	0	27	3	K.TLPEIFK.T
84		497.2492	992.4839	992.5001	-0.0162	0	25	1	K.YLDPLMKN.L
87		505.2698	1008.5251	1008.495	0.0301	0	(17)	5	K.YLDPLMKN.L + Oxidation (M)
110		556.3372	1110.6598	1110.6648	-0.005	0	38	1	K.LGPETVAALK.V
155		629.0162	1257.6179	1257.6241	-0.0061	0	14	1	K.FTVNTGISYK.E
166		642.8271	1283.6396	1283.655	-0.0154	0	13	10	K.VYGFNAPFLK.T
172		661.3454	1320.6762	1320.6925	-0.0163	0	35	1	K.ALYELLSAADQK.A
189		677.8231	1353.6316	1353.6347	-0.003	0	27	1	R.ANTFENSTNARA
235		761.3721	1520.7297	1520.747	-0.0173	0	12	1	K.VLNEQVDSYGLR.E
248		587.6303	1759.8691	1759.874	-0.0049	0	29	1	K.LYQENSPLYTNETR.L

3: Here we are filtering out any protein matches that contain any peptide with any variable modification.

It would, of course, be possible to generate a filter which acted on any, or all, of these criteria in one go, so it offers a very flexible and powerful way of mining larger Mascot search results

Help

Online context sensitive help

Designing an Excel Query

To create an Excel SQL Query, you must first define the database connection (see [Connecting to Mascot Integra using ODBC](#)) and then specify which view and columns you wish to work with. Clauses can be added to the query to filter the results returned and/or display only those results meeting the conditions of the 'where' clause. Once the query has been tested and saved, it must be imported into Excel. Specific formatting can be added to the worksheet before it is saved in Excel. The saved Excel file is then imported into Mascot Integra and saved as an **Excel Report Template**.

To add a new Excel query:

1. From the Site Map, click **New Excel SQL Query** to open the Excel Query definition wizard.
2. Step 1: Enter the **ODBC** details. Ensure that the Report view owner is set to the name of the database. Click the **Next** button.
3. Step 2: Select the views that you want to include in the query. Select one or more views in the 'Available views' list and click the **Add** button. To remove a view from the 'Selected views' list, select it and click the **Remove** button. Click [here](#) for details of the Available Views.
4. Click the **Next** button to continue.
5. Step 3: Select the columns that you want to include and click the **Add** button. The list of 'Available columns' is based on the views that you selected in the previous step. To re-order the columns in the 'Selected column' list, select a column and use the **Up** and **Down** buttons to move it up or down in the list.
6. Click the **Next** button to continue.
7. Step 4: Specify any 'where' clauses to limit the data returned. To add a clause, click the **Add new clause** button. Choose the Clause type, Column name, Operator and Value. If you have the value set to '?', the user will be prompted to enter a value at run-time (except for the LIKE clause queries which cannot take run-time parameters). To remove a clause, click the corresponding **Remove Clause** button.
8. Click the **Next** button to continue.
9. Step 5: Select values that were specified in the previous step to test the query. The separate clauses will be automatically joined with an AND clause. This can be changed to OR, if required.
10. Click the **Next** button to continue.
11. Step 6: If you need to edit the query, click in the testing box and make the required changes. Alternatively, copy and paste the query into a text editor. If you need to make extensive changes, you will need to be familiar with SQL to make changes to the query.
12. Step 7: Click the **Next** button to test the query. If the query is successful, the first 10 rows returned will be shown. If the query does not return the correct information, click the **Back** button and make the required changes.
13. Click the **Next** button to continue.
14. Step 8: Enter a name for the query. Spaces are allowed in the name but do not include a file extension since this will be treated as part of the query name. A file extension (.dq3) is automatically added when the query is saved.
15. Click the **Next** button.
16. Review the details on this page. The query details and query name are shown.
17. Step 9: Click the **Next** button to confirm and save the query. The query will be saved in a shared folder on the server called `Integra\ExcelQuery`.

MASCOT : Problem solving with Mascot Integra

© 2006 Matrix Science



Don't worry if some of that looked complicated, online context sensitive help is at hand.

You can access help on each of the menu options from the tramline page or open the relevant help section directly from your current location in Integra.

Shown here is a walk through for designing new excel queries

Acknowledgments

Pascal Wather

- Department of Biochemistry, University of Zurich, Switzerland

and Bertran Gerrits for mouse ICAT dataset

- Functional Genomics Center Zurich, Uni|ETH Zurich, Switzerland

Katheryn Resing for A8 dataset

- University of Colorado, USA

I would like to thank two labs for the data sets, Pascal Wather and Bertran Gerrits for the use of the Mouse ICAT data and Katheryn Resing for making the A8 dataset publicly available on the Peptide Atlas website.