

MASCOT*Integra*

*Data management for
Proteomics*

ABRF 2005



Mascot Integra: Data management for proteomics

- What is Mascot Integra?
- What Mascot Integra is not
- Security and Electronic signatures in Mascot Integra
- Instrument integration in Mascot Integra
- Designing and running proteomics experiments
- Results and reporting in Mascot Integra.

ABRF 2005



Mascot Integra is our new solution for proteomics sample and data management.

During this talk, I will be discussing these topics

What is Mascot Integra?

- Fully functional 'out-the-box' solution for proteomics workflow and data management
- Support for all the major mass-spectrometry data systems
- Powered by the Sapphire™ LIMS package from LabVantage Solutions Inc
- Oracle 9 database
- Scalable to the largest projects.

ABRF 2005



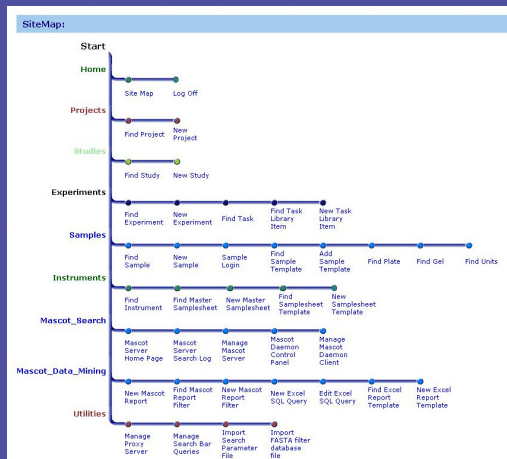
Mascot Integra is supplied as a ready to run system. It does not require the extensive setup and customisation associated with a traditional LIMS package.

Rather than re-invent the wheel, we have partnered with LabVantage Solutions Inc, (www.lims.com). Their Sapphire LIMS package provides the sample tracking and workflow modelling functionality for Mascot Integra

Using the Oracle database management system enables the database to scale efficiently as your data management requirements grow

What is Mascot Integra?

- **Laboratory Information Management for**
 - Sample tracking
 - Isolation / fractionation / purification
 - Splitting / combining
 - 1D and 2D Chromatography
 - 1D and 2D Electrophoresis
 - Image analysis
 - Spot picking
 - Digestion / derivatisation
 - MS and MS/MS analysis
 - MS data reduction
 - Mascot database search
 - Data warehouse
 - Result reporting
 - Data mining



ABRF 2005



We aim to support all aspects of the workflows associated with Proteomics laboratory processes.

What Mascot Integra is not

- **An enterprise LIMS**
 - Does not require a customised configuration or extensive consultancy
 - Won't run your entire organisation
 - No inventory management or reagent re-ordering
 - Not a substitute for instrument data systems
- **As expensive as an enterprise LIMS.**

ABRF 2005

{MATRIX}
{SCIENCE}

Mascot Integra will supply the LIMS functionality required to run a proteomics facility. It is not an "enterprise" LIMS system, and does not contain functionality for (say) genomics experiments

LabVantage Solutions Inc

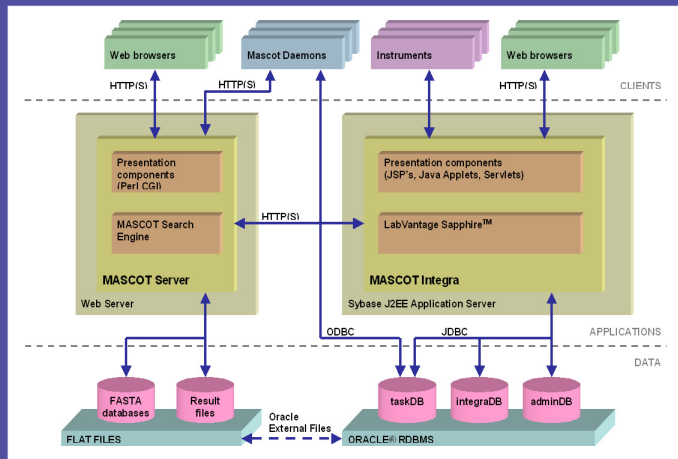
- **LabVantage LIMS expertise**
 - 24 year track record
 - Scale up to LabVantage Enterprise LIMS products
 - Call on World-Wide Professional Support Services

ABRF 2005



Mascot Integra architecture

- 3 tier system
 - Oracle database server
 - Sybase Enterprise Application Server running a J2EE web application
 - All user functionality available through Internet Explorer



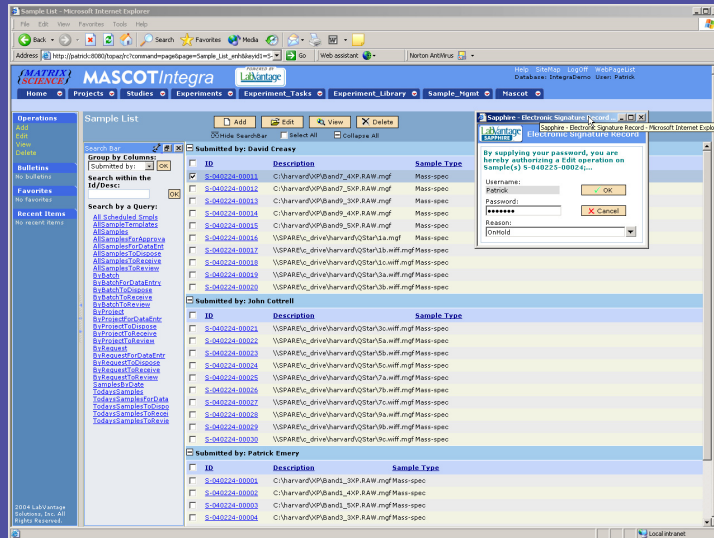
ABRF 2005

**MATRIX
SCIENCE**

All Mascot Integra functionality is accessible through a standard web browser.

CFR21 part 11 ERES compliant

- Electronic signatures
- Maintains audit trail
- Role and status based user access
- Automatic inactivity logout (definable period)



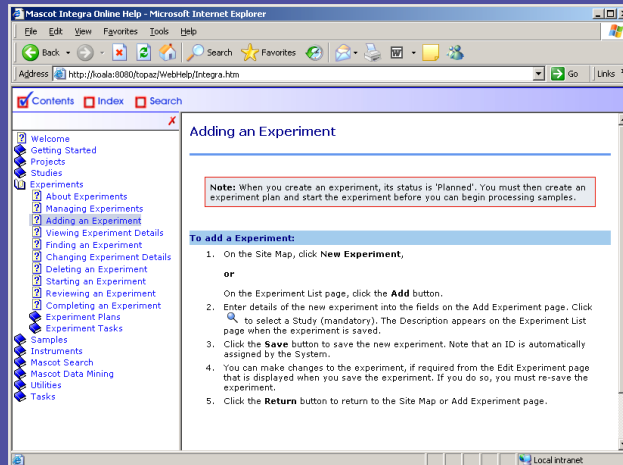
ABRF 2005



Mascot Integra complies with the requirements of FDA CFR21 part 11 ERES

Online help

- Extensive context sensitive help available on web-based help system



ABRF 2005



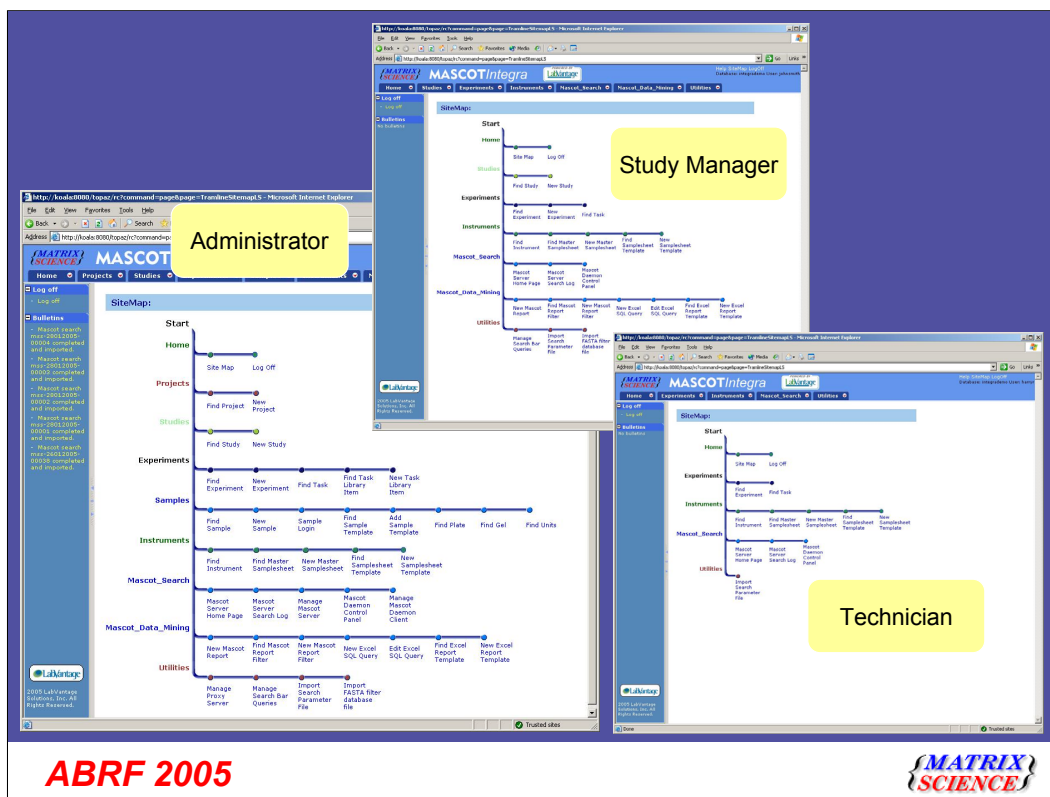
Security in Mascot Integra

- Role and project/study based security
- Assign roles to users
- Page access limited by user role
- Assign users to projects and studies
- Access to experiments limited by study membership.

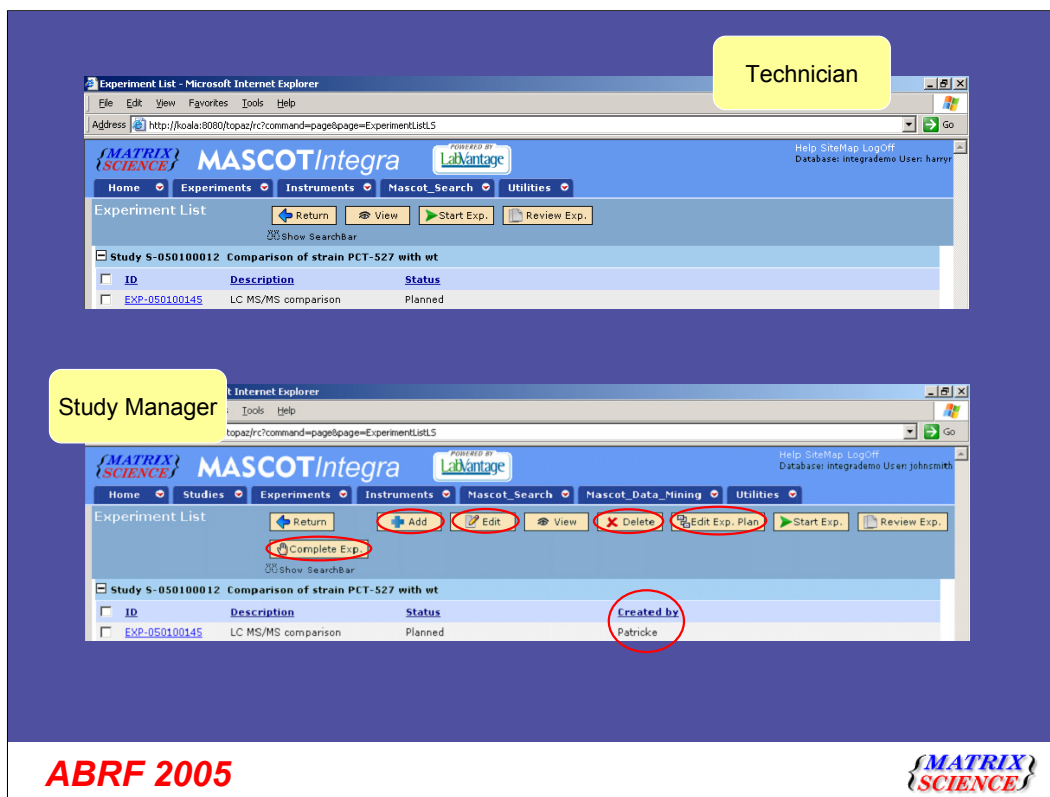
ABRF 2005



Users can be assigned to multiple projects/studies and assigned multiple roles (e.g. a user can be both a Study Manager and an Analyst)



To illustrate this, the administrator has full access to the system and sees all the possible links from the Mascot Integra site-map. The study manager has more limited access to the system (for example, the project links are not visible to them). The technician sees only the links required to perform specific tasks.



In addition to limiting access to complete pages, individual controls on a page can be altered according to the user's role. Here we have the same page (a list of available experiments) as viewed by a technician and a study manager.

Instrument Integration

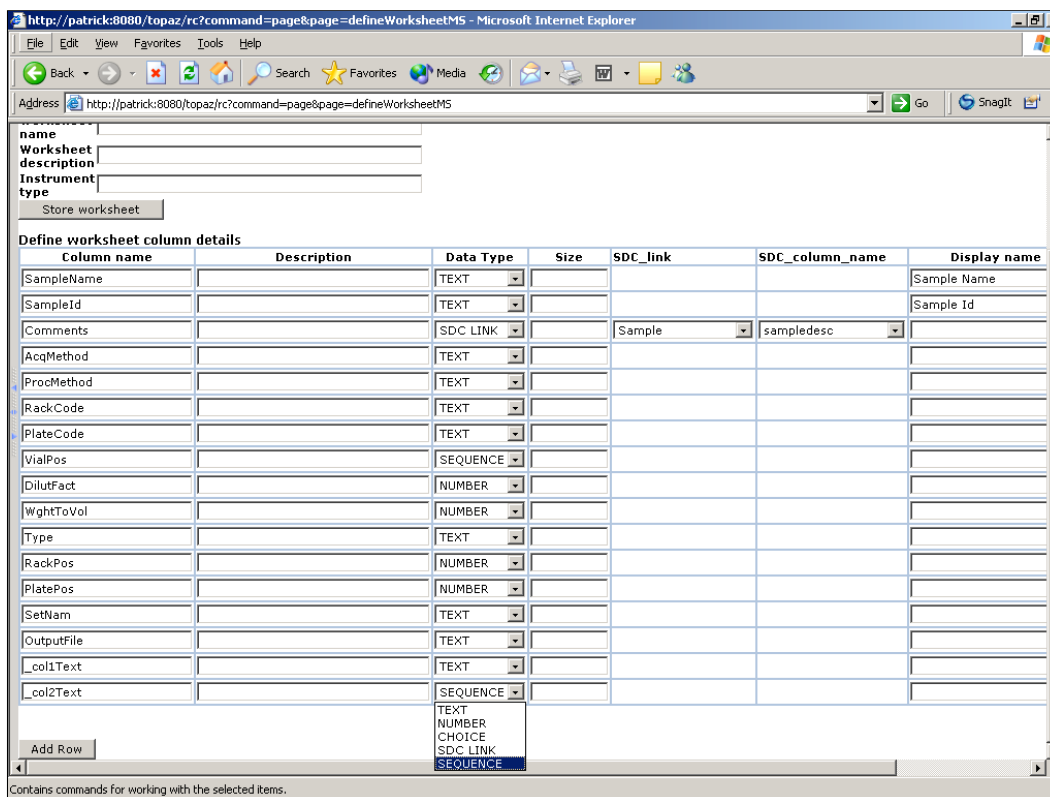
- Not “running” the instrument from the LIMS
- Handled via Sample / work sheet exchange
- Comes with a range of master sample sheets for the main instrument types, data systems and manufacturers
- Flexible – you can design your own master sample sheets and sample sheet templates
- Output CSV, TXT (tab delimited) and Microsoft Excel files.

ABRF 2005

**MATRIX
SCIENCE**

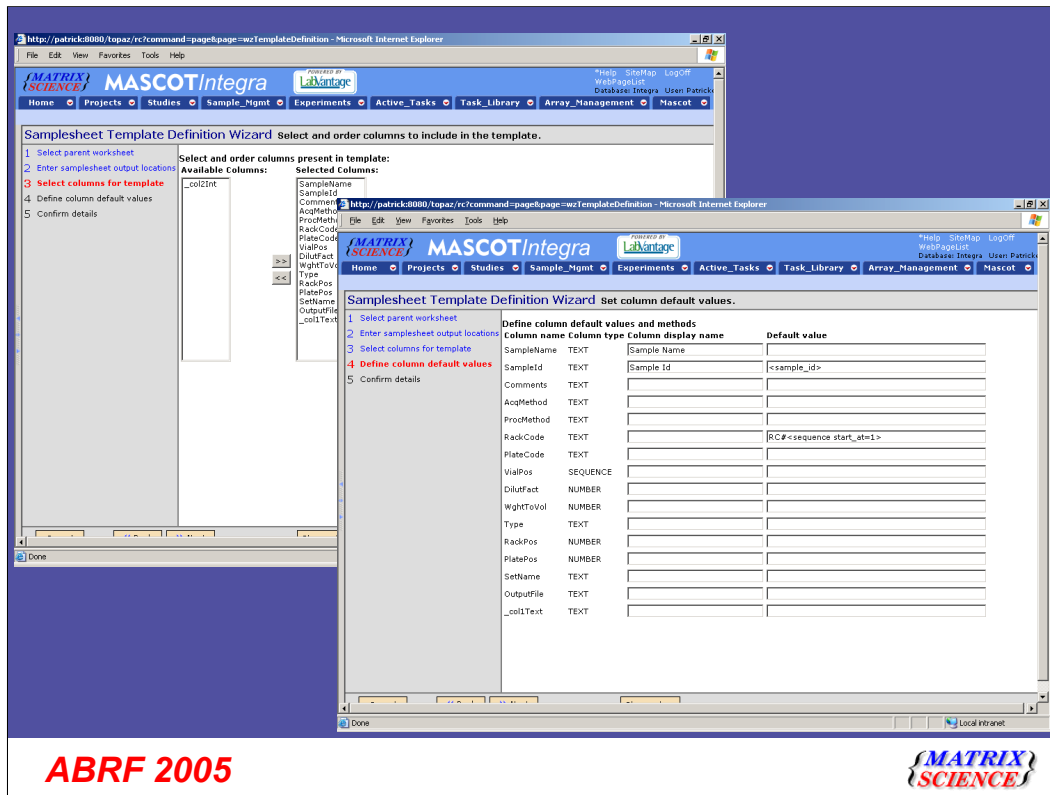
Instrument integration is handled via sample sheet exchange.

Our aim is to minimise the requirement for copying data from the output of one instrument data system to the input of another. This reduces the time involved in setting up instrument runs and the opportunity for errors.



For a new instrument, there is a one-time setup to design a master sample sheet containing all the possible data elements supported by the instrument data system.

In addition to standard data types (text, numeric etc) you can specify a column as an SDC link. This means that the column defines a live link back to data held in a field in the Mascot Integra database. The data retrieved can be determined by the sample identifier associated with the row of the sample sheet.



Once the master sample sheet has been defined, a subset of the columns is selected for a sample sheet template.

At this stage, you can specify default values for the columns

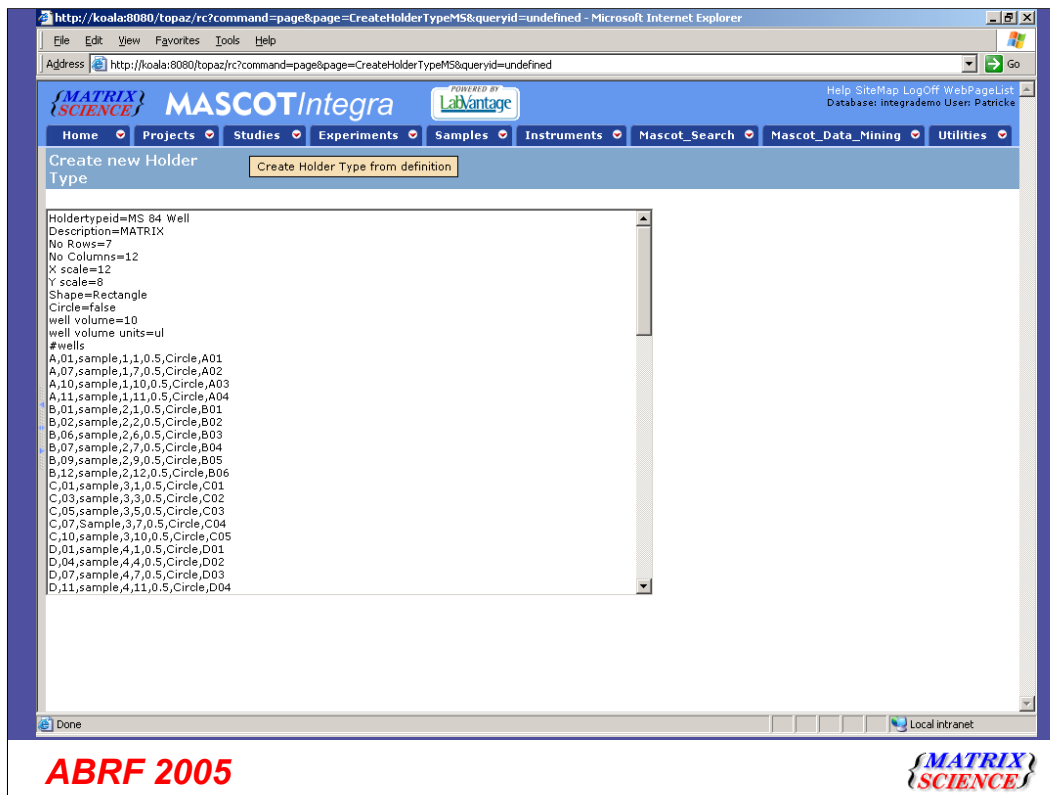
Custom Plate design

- Ships with a range of common plate types
 - 96 Well, Mass-Lock MALDI plates etc
- Custom definition file enables user to define new plate types
 - Flexible

ABRF 2005

**MATRIX
SCIENCE**

A wide variety of MALDI target plates are used by different instruments. To accommodate additional or specialist plate types, Mascot Integra even allows you to define new plate types.



The format is self explanatory. The definition on this slide produced the plate shown in the next slide...

http://koala:8080/topaz/rc?command=page&page=AdhocPlateLoadL5&queryid=AllSamples - Microsoft Internet Explorer

Address http://koala:8080/topaz/rc?command=page&page=AdhocPlateLoadL5&queryid=AllSamples

MATRIX SCIENCE **MASCOTIntegra** POWERED BY LabVantage Help SiteMap LogOff WebPageList Database: Integrademio User: Patrickc

Home Projects Studies Experiments Samples Instruments Mascot_Search Mascot_Data_Mining Utilities

Transfer Volume: Transfer All Content Specify Transfer Volume ul Undo Done Cancel

Select Method: List Page Transfer Style: Horizontal Vertical Block

Source List: Sample SDC Target Plate:

<input type="checkbox"/>	Sample ID	Description	Total Content	Units
<input type="checkbox"/>	S-050126-00001	Saliva	89	ul
<input type="checkbox"/>	S-050129-00001	Trypsin Digest	50	ul
<input type="checkbox"/>	S-050129-00002	wt S.pombe culture	0	ul
<input type="checkbox"/>	S-050129-00003	5x10 ⁹ S.pombe wt cells resuspended in 1ml grinding/resuspension buffer	0	ul
<input type="checkbox"/>	S-050129-00004	S.pombe lysate resuspended in 2D buffer	600	ul
<input type="checkbox"/>	S-050129-00005		0	ul
<input type="checkbox"/>	S-050129-00006		48	ul
<input type="checkbox"/>	S-050129-00007	Whole cell protein extract 6.5mg/ml concentration	130	ul
<input type="checkbox"/>	S-050129-00008	Whole cell protein extract 6.5mg/ml concentration	150	ul
<input type="checkbox"/>	S-050129-00009	Whole cell protein extract 6.5mg/ml concentration	150	ul
<input type="checkbox"/>	S-050129-00010	1l S.pombe cells	0	ul
<input type="checkbox"/>	S-050129-00011	S.pombe culture resuspended in 1ml grinding/resuspension buffer	0	ul
<input type="checkbox"/>	S-050129-00012		0	ul
<input type="checkbox"/>	S-050129-00013	Purified S.pombe whole cell extract	49.5	ul
<input type="checkbox"/>	S-050129-00014	Purified S.pombe whole cell extract	139.5	ul
<input type="checkbox"/>	S-050129-00015	Purified S.pombe whole cell extract	149.5	ul
<input type="checkbox"/>	S-050129-00016	Purified S.pombe whole cell extract	149.5	ul
<input type="checkbox"/>	S-050129-00017		2	mm

New Open Save

New

Expt List Save List Print List

Target Plate Log:

No rows found.

Done Local intranet

;-)

Experiment design:

- Experiments are broken into a series of experimental tasks
- Experimental tasks have inputs and outputs (e.g. Sample, Plate)
- Experimental tasks can be linked together based on their inputs and outputs
- Each task is then associated with a series of pages which model and setup the physical task
 - While running a task, the user can add task and experiment notes (free text) at any time
- A comprehensive library of flexible proteomics tasks is provided.

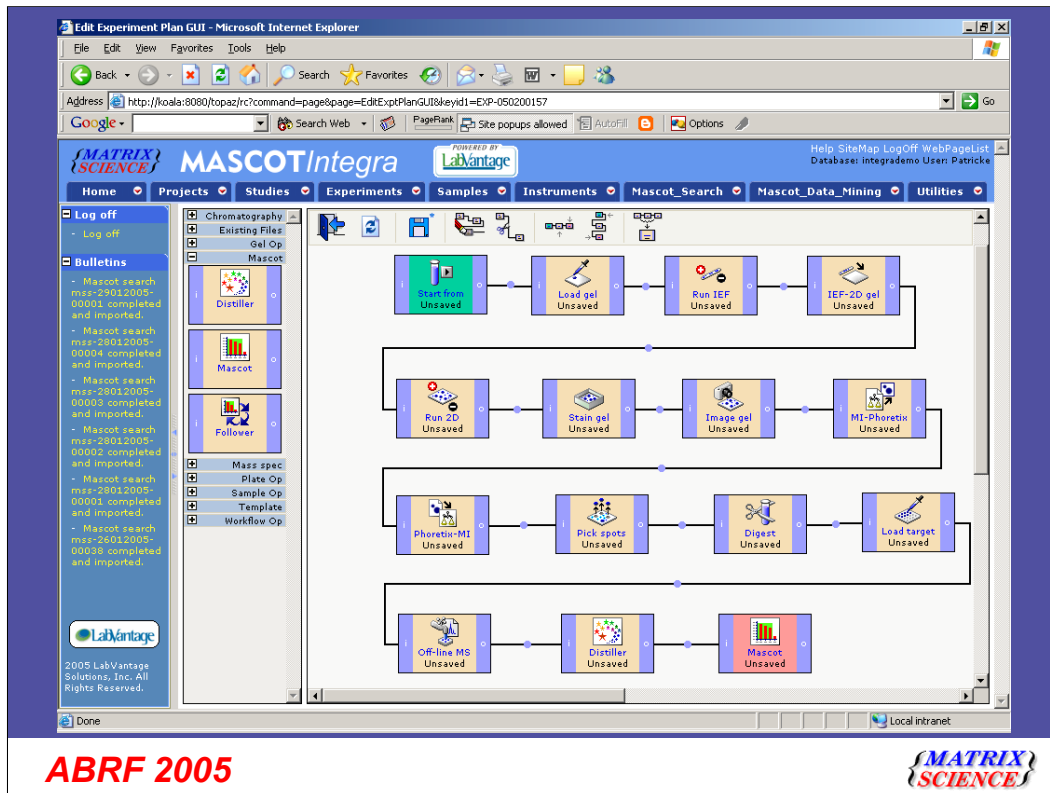
ABRF 2005

**MATRIX
SCIENCE**

Individual tasks in an experiment represent discrete laboratory processes e.g. running a 2D gel, loading a 96 well plate, enzyme digestion etc.

For the experimental task of loading a MALDI target plate, the input would be a sample list, and the output would be a loaded target.

Experimental tasks can have multiple inputs and outputs.



ABRF 2005

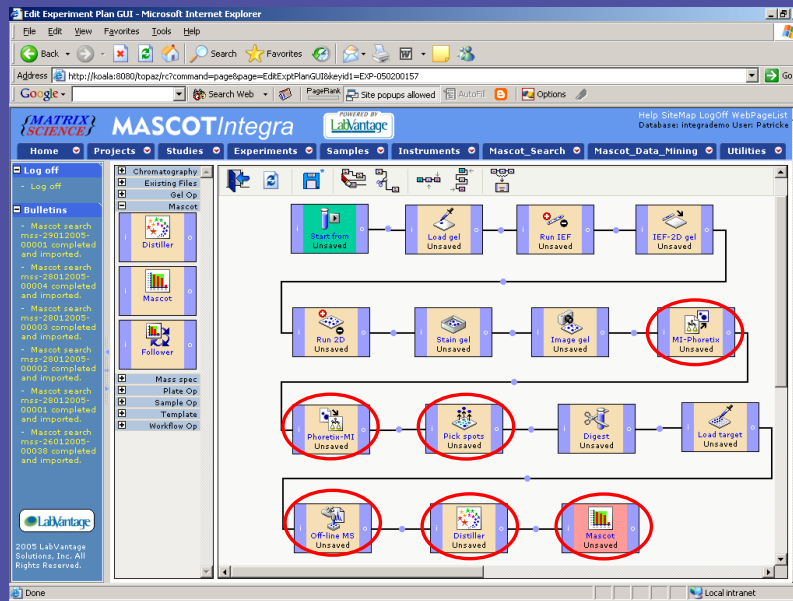


A complete workflow can be saved as a new experimental task, which is available from the task list menu bar. Any default values which were entered against the experimental tasks would also be saved in this new template. This provides a simple method for setting up the default workflows for your laboratory.

Running experiments:

- Now we'll go through setting up and running a common type of proteomics experiment
 - 2D Gel analysis
 - Image analysis with Nonlinear Dynamics Phoretix™ Software
 - MS run
 - Defining Raw data processing with Mascot Distiller
 - Setting up a batch search with Mascot Daemon
 - Running and importing Mascot searches
 - Examining the results

Simple 2D Gel analysis workflow



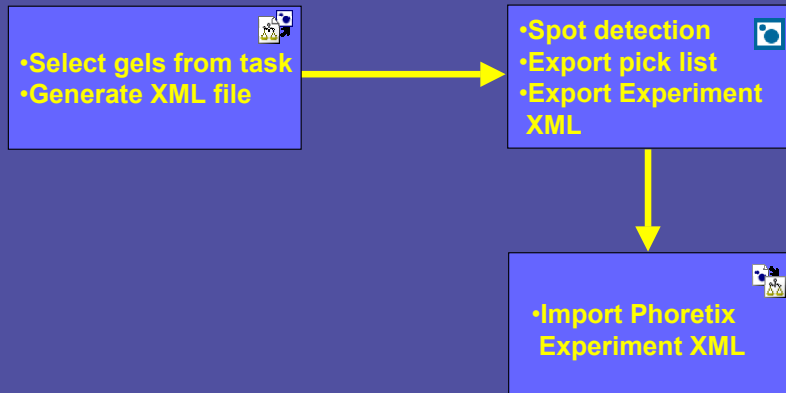
ABRF 2005



Rather than running through the entire work flow, we'll just be looking at the circled tasks. These define the Mascot Integra interfaces to the NonLinear Dynamics Phoretix package, to the mass-spec data system, and to the Mascot search engine.

Nonlinear Dynamics Phoretix integration

- Handled via Phoretix XML exchange format



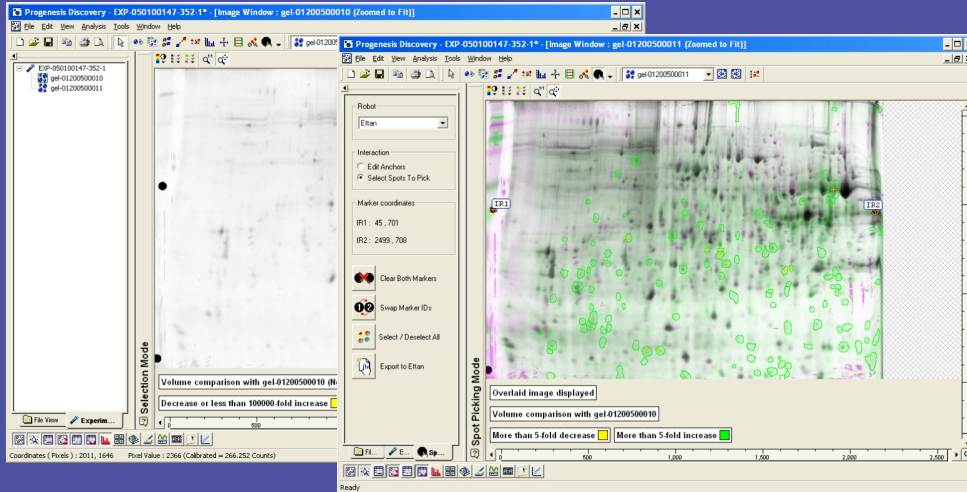
ABRF 2005

**MATRIX
SCIENCE**

In this workflow, we are tracking the loading of the gels with samples, and the capture of the stained gel images. The information is transferred between Mascot Integra and Phoretix using the Phoretix XML Exchange Format.

Spot detection and picklist generation (for spot cutting robots) is handled by Phoretix. This information is the imported back into Mascot Integra.

Nonlinear Dynamics Phoretix integration 2



ABRF 2005



Nonlinear Dynamics Phoretix integration 3

- Import the Phoretix experiment XML export file into Integra

Task: Spot Picking

Deselect All Collapse All

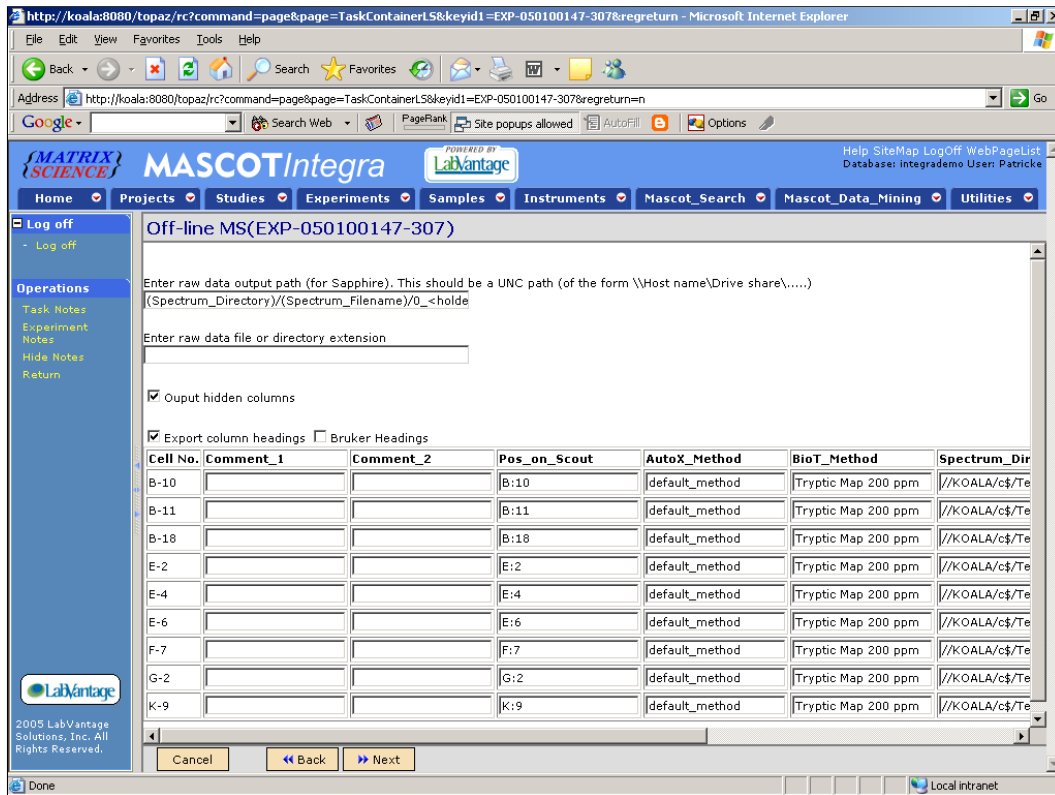
Scan in Ids:

<input checked="" type="checkbox"/>	ID	Description	Source Spot Number	X Position	Y Position
<input checked="" type="checkbox"/>	gs-012005-00327		1524	2235	505
<input checked="" type="checkbox"/>	gs-012005-00328		2190	912	850
<input checked="" type="checkbox"/>	gs-012005-00329		2319	1410	939
<input checked="" type="checkbox"/>	gs-012005-00330		2333	1522	946
<input checked="" type="checkbox"/>	gs-012005-00331		2356	1492	969
<input checked="" type="checkbox"/>	gs-012005-00332		2435	1552	1021
<input checked="" type="checkbox"/>	gs-012005-00333		2471	1952	1053
<input checked="" type="checkbox"/>	gs-012005-00334		2481	1920	1061
<input checked="" type="checkbox"/>	gs-012005-00335		2977	1116	1574
<input checked="" type="checkbox"/>	gs-012005-00336		2979	1030	1577

ABRF 2005

**MATRIX
SCIENCE**

We can choose to import all the detected spots or just those spots which we have chosen for picking.



Next, we select a pre-defined template for a Bruker TOF instrument samplesheet

Some values are automatically filled in from the database

Export as CSV or xls format file for import into Bruker's FlexControl data system

Peak detection with Mascot Distiller

- Mascot distiller used for automated peak detection
- Used in conjunction with Mascot Daemon for automated search submission
- Import of peak lists created by other software is supported.

ABRF 2005



The screenshot displays the Mascot Distiller software interface. On the left, a menu is open with options like 'New Project', 'Open Project...', 'Close Project', 'Save Project', 'Print...', and 'Exit'. The main window shows search results for 'myoglobin tryptic digest' with a list of peptides and their scores. Two chromatograms are visible: a Total Ion Chromatogram (TIC) and a Summed Scans plot. A yellow callout box in the upper right corner states 'Easy to use, cross platform data browser'. At the bottom left, the text 'ABRF 2005' is displayed, and at the bottom right, the 'MATRIX SCIENCE' logo is present.

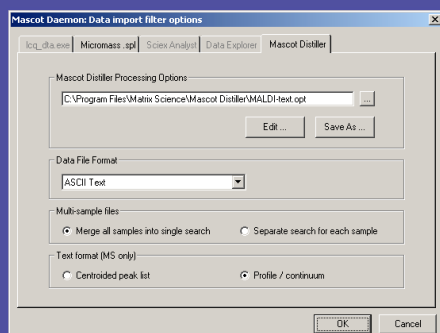
Mascot Distiller offers optimised peak detection for all the popular raw data formats.

Peak detection with Mascot distiller continued...

Distiller(EXP-050100147-308) Define sample raw data reduction

Select raw data processing method:

Mascot Distiller online with Mascot Daemon
Mascot Distiller online with Mascot Daemon
Import peaklist file



Distiller(EXP-050100147-308) Define sample raw data reduction

Select Mascot distiller options file:

XIMASS-XTOF_MS.opt

Select data file format

Bruker XIMASS/XTOF

ABRF 2005



The Mascot Distiller library is used as a data import filter by Mascot Daemon

Mascot search submission

- Automated search submission using Mascot Daemon
- Multiple Mascot daemon clients running as services
- Parameter sets are held centrally in the Mascot Integra database, not in local parameter files.

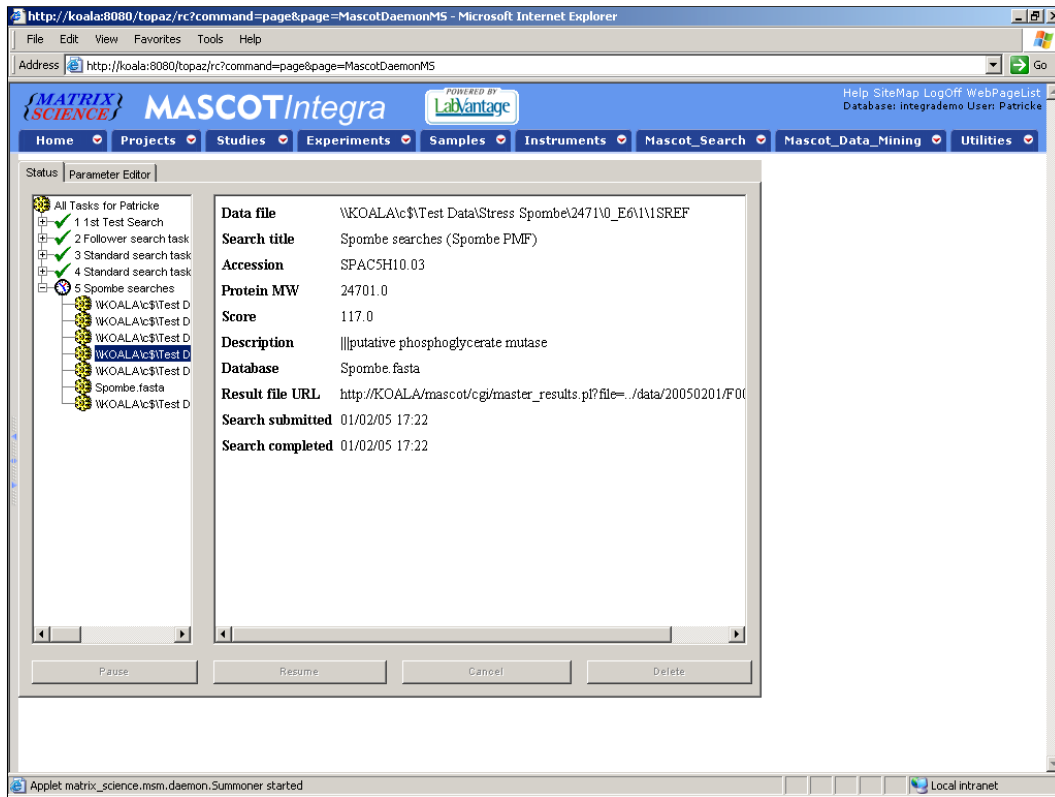
ABRF 2005



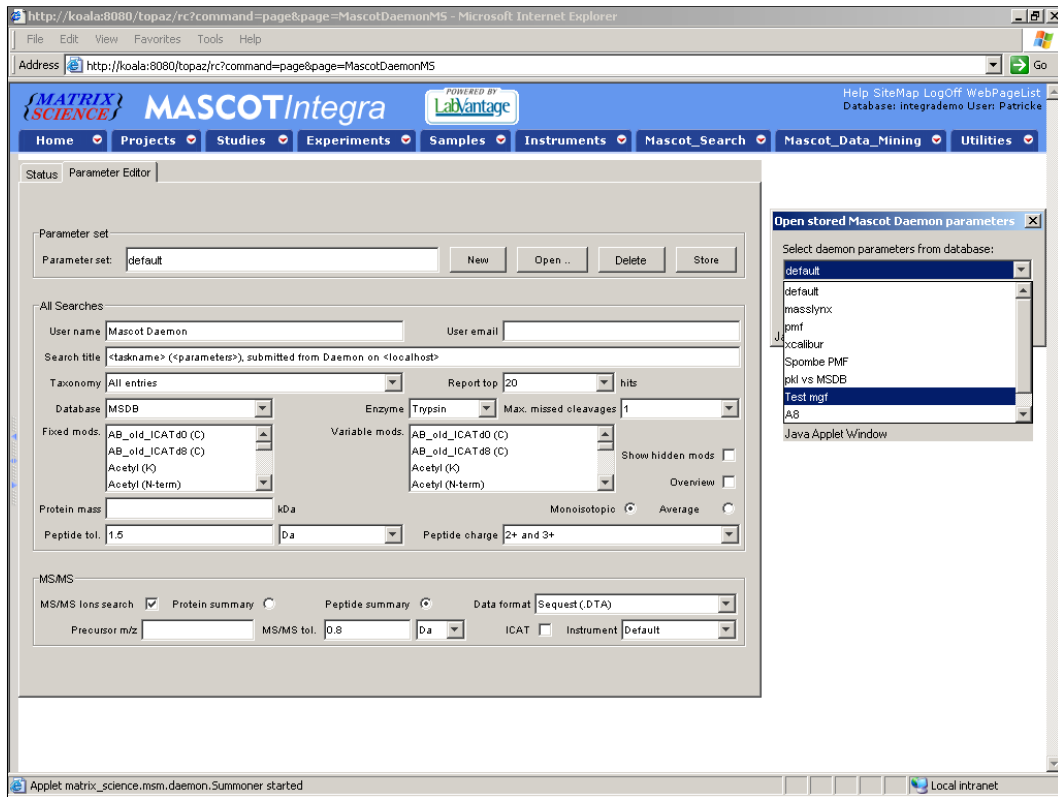
Mascot search submission is handled by Mascot Daemon.

A common arrangement is to have a Mascot Daemon client on each instrument data system, all communicating with the Mascot Integra database. Users assign processing and search tasks to the relevant Mascot Daemon client via the Mascot Integra interface.

Parameter sets are held in the central database tables, ensuring that the parameter sets used by all the Mascot Daemon clients are identical.



This looks like Mascot Daemon, but it is actually a Java Applet hosted in the web browser. This can be used for monitoring and managing all of the search tasks on all of the Mascot Daemon clients.



The Java applet is also used to manage the central sets of search parameters

The screenshot shows the Mascot Integra web application interface. At the top, there is a navigation menu with options: Home, Projects, Studies, Experiments, Samples, Instruments, Mascot_Search, Mascot_Data_Mining, and Utilities. Below the menu is the 'Mascot Search' section, which contains a table of search results.

Search filename	Sample description	Daemon Task Id	Search status	Comments
\\KOALA\c\$\Test Data\Stress Spombe\2333\0_B10\1\1SREF		5	Awaiting Mascot results	Submitted file details to Mascot Daemon
\\KOALA\c\$\Test Data\Stress Spombe\2356\0_B11\1\1SREF		5	Awaiting Mascot results	Submitted file details to Mascot Daemon
\\KOALA\c\$\Test Data\Stress Spombe\2435\0_B18\1\1SREF		5	Finished	Finished importing protein hits
\\KOALA\c\$\Test Data\Stress Spombe\2190\0_E2\1\1SREF		5	Finished	Finished importing protein hits
\\KOALA\c\$\Test Data\Stress Spombe\2319\0_E4\1\1SREF		5	Parsing result file	Parsing initiated
\\KOALA\c\$\Test Data\Stress Spombe\2471\0_E6\1\1SREF		5	Queing results for importing	
\\KOALA\c\$\Test Data\Stress Spombe\2481\0_F7\1\1SREF		5	Queing results for importing	
\\KOALA\c\$\Test Data\Stress Spombe\2977\0_G2\1\1SREF		5	Awaiting Mascot results	Submitted file details to Mascot Daemon
\\KOALA\c\$\Test Data\Stress Spombe\1524\0_K9\1\1SREF		5	Awaiting Mascot results	Submitted file details to Mascot Daemon

Below the table, there is a note: "If you have entered incorrect details for any search file, then pause the Mascot Daemon task and you will be offered a link to edit the file details for that task. Then either restart the Mascot Daemon service or use the Mascot Daemon applet to cancel the task click the refresh button to resubmit the searches. The resubmitted searches will use the edited file paths." Below this note is a link "View Mascot Daemon Applet" and a statement "This page will automatically refresh every 60 seconds." There is also a "Refresh" button and a warning: "Do not use the browser refresh button as you will loose any task and experiment notes you have added."

The footer of the page contains "ABRF 2005" on the left and the Matrix Science logo on the right.

Results are automatically parsed and imported into Mascot Integra as each search completes.

1 Selected hit: [gi|12876711](#) ALS44231 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DI019YF12 5 prime

Check to validate and import peptide matches into Integra database

8 good peptide matches.

Comments:

[gi|12876711](#) Mass: 37305 Total Score: 953 Peptides Matched: 19
 ALS44231 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DI019YF12 5 prime

Approve Match?	Query	Observed Mr(exp)	Mr(Calc)	Delta	Miss	Score	Rank	Peptide	
<input type="checkbox"/>	11	415.19	828.36	828.5	-0.13	0	31	2	VLDLELK
<input type="checkbox"/>	12	415.19	828.37	828.51	-0.14	0	36	2	NALLSLAK
<input type="checkbox"/>	32	540.65	1079.29	1079.48	-0.19	0	20	2	SEIDMNDIK + Oxidation (M)
<input checked="" type="checkbox"/>	45	607.16	1212.32	1212.53	-0.21	0	68	1	DITSDTSGDFR
<input checked="" type="checkbox"/>	53	631.7	1261.38	1261.59	-0.22	0	70	1	TPAQFDADELK
<input type="checkbox"/>	65	457.85	1370.54	1370.77	-0.23	1	43	1	VLDLELKGDIK
<input checked="" type="checkbox"/>	69	694.25	1386.49	1386.76	-0.27	0	73	1	GVDEATIIDLTK
<input type="checkbox"/>	91	515.2	1542.58	1542.86	-0.28	1	45	1	GVDEATIIDLTKR
<input type="checkbox"/>	92	772.3	1542.58	1542.86	-0.28	1	(6)	8	GVDEATIIDLTKR
<input checked="" type="checkbox"/>	93	775.76	1549.5	1549.81	-0.31	0	68	1	GTDVNVFNTILTR
<input type="checkbox"/>	98	547.49	1639.45	1639.77	-0.32	1	(41)	1	DLAKDITSDTSGDFR
<input type="checkbox"/>	99	820.75	1639.48	1639.77	-0.29	1	58	1	DLAKDITSDTSGDFR
<input checked="" type="checkbox"/>	103	851.77	1701.52	1701.88	-0.36	0	103	1	GLGTDIEDTLIEILASR
<input checked="" type="checkbox"/>	105	870.21	1738.41	1738.73	-0.32	0	103	1	SEDFGVNEDLADSDAR
<input type="checkbox"/>	111	592.87	1775.6	1775.93	-0.33	1	16	1	AAYLQETGKPLDETLK
<input type="checkbox"/>	123	476.92	1903.67	1904.03	-0.36	2	28	1	AAYLQETGKPLDETLK
<input type="checkbox"/>	131	707.22	2118.63	2119.08	-0.45	1	40	1	AAMKGLGTDIEDTLIEILASR + Oxidation (M)
<input checked="" type="checkbox"/>	133	1070.83	2139.64	2140.01	-0.37	0	82	1	QAWFIENEQEYVQTVK

All the standard Mascot reports can be generated from the data held in the database. In addition, individual protein and peptide matches can be annotated and approved.

http://koala:8080/topaz/rc?command=page&page=ProteinViewM5 - Microsoft Internet Explorer

Address http://koala:8080/topaz/rc?command=page&page=ProteinViewM5

gii12876711 has previously been selected by:
 Patricke on 03-Feb-2005 14:18:24. 8 good peptide matches.

Translated in frame

Nominal mass (Mr): **37305.44**; Calculated pI value: **8.33**
 NCBI BLAST search of [gii12876711](#) against nr
 Unformatted [sequence string](#) for pasting into other applications
 Variable modifications: Carbamidomethyl (C), Oxidation (M), Propionamide (C)

Cleavage by Trypsin/P: cuts C-term side of KR
 Total number of queries searched: 19
 Total number of queries matched: 19
 Sequence Coverage: 50.3%

Matched peptides are shown in **bold red**
 Peptides which have been previously selected are highlighted with a blue background

1 FSSLQEGRDK DTFSSKMANVS EFLK**QAWFIE NEEQEYVQTV KSSKGGGSA**
 51 **VSPYPTFNPS SDVAALHKAI** MVR**GVDEATI IDILTKR**NNA QRQIKAAYL
 101 **QETGKPLDET LKKA**LTGHLE YEVVLALLK**TP AQFDADELRA** AMKGLGTDED
 151 **TLIEILASRT NKEIR**DIRV YREELK**RDLA KDITSDTSGD** FRNALLSLAK
 201 GDRSEDFGVN EDLADSDARA LYEAAGER**RG TDVNVFNTIL** TTSYFPQLRR
 251 VFQKYTKYSK HDMMK**VLDE LKGDIEK**CLT AIVKCAQANQ LSLQRSFIKP
 301 **KVLELAIRH** _SGLWFFV**LK** LT_MISKHSI RRVW

Click [here](#) to view a printer friendly version of this page.

Show predicted peptides

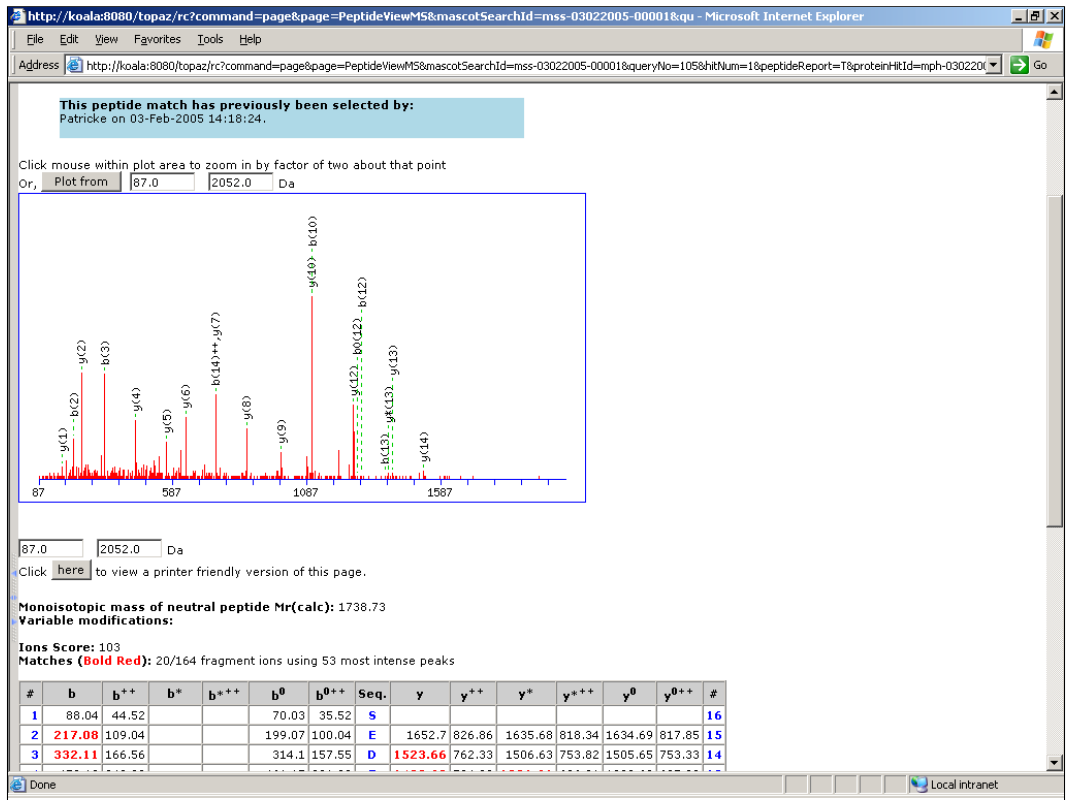
Sort Peptides By Residue Number Increasing Mass Decreasing Mass

Start	End	Observed Mr(exp)	Mr(calc)	Delta	Miss	Sequence	Ions score
25	41	1070.83	2139.64	2140.01	-0.37	0 QAWFIENEQEYVQTVK	(82.64)
45	68	785.91	2354.72	2355.15	-0.43	0 GGPGSAVSPYPTFNPSDVAALHK	(65.41)
74	86	694.25	1386.49	1386.76	-0.27	0 GVDEATIIDILTK	(73.06)
74	87	515.2	1542.58	1542.86	-0.28	1 GVDEATIIDILTKR	(45.01)
97	112	592.87	1775.6	1775.93	-0.33	1 AAYLQETGKPLDETLK	(16.58)
97	113	476.92	1903.67	1904.03	-0.36	2 AAYLQETGKPLDETLKK	(28.1)
129	139	631.7	1261.38	1261.59	-0.22	0 TPAQFDADELK	(70.99)
140	159	707.22	2118.63	2119.08	-0.45	1 AAMKGLGTDEDTLIEILASR + Oxidation (M)	(40.74)

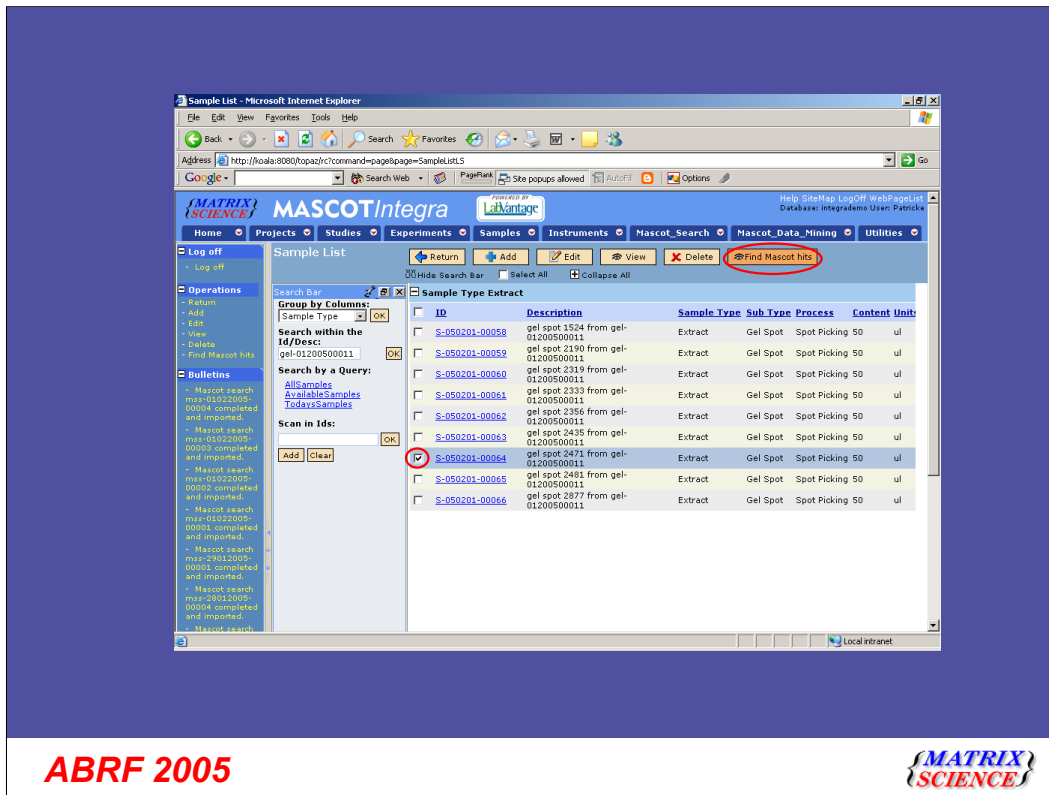
Done Local intranet

Approval information is flagged up on the reports, such as Protein View

...



... And Peptide View



ABRF 2005



One advantage of tracking workflows in Mascot Integra is that all of the relationships between parent and child samples are available. Here, we drill down from a gel spot sample...

The screenshot displays the MascotIntegra web application interface. The browser window title is "Protein hit list - Microsoft Internet Explorer". The address bar shows the URL: <http://koala:8080/topaz/c?command=page&page=ProteinListM5&keyId=mss-01022005-00007>. The page header includes the "MATRIX SCIENCE" logo and "MASCOTIntegra" branding. A navigation menu contains items like Home, Projects, Studies, Experiments, Samples, Instruments, Mascot_Search, Mascot_Data_Mining, and Utilities. The main content area is titled "Protein hit list" and shows search results for "Mascot Search Id: mss-01022005-00007". A search bar and "Group by Columns" options are visible. A table of results is displayed with the following data:

ID	Accession	Description	Hit Rank	Mascot Protein Score
msh-01022005-000124	SPBC1815.01	leno1 lenolase	1	135

At the bottom of the slide, the text "ABRF 2005" is displayed on the left and the "MATRIX SCIENCE" logo on the right.

... to the corresponding mascot protein hit

Filtering reports

- Flexible report filters
- Ships with a range of useful filters
 - E.g. limit returned results to significant matches
- User definable
- Requires some knowledge of SQL.

ABRF 2005

{MATRIX}
{SCIENCE}

Because all of the results data is held in relational database tables, we can define flexible filters to select results for a report.

http://141-dmc:8080/topaz/r7command-page8page-MasterResultsMS Microsoft Internet Explorer

Address http://141-dmc:8080/topaz/r7command-page8page-MasterResultsMS

Display filter options:
SP or TP

Selection filter options:
Include keywords:
Exclude keywords:

Expectation value threshold: 0.05
Submit

Click [here](#) for a printer friendly version of this page.

Peptide summary report
14 Selected hit: *IP100013769 Tax_id=9606 Alpha enolase, lung specific*
 Check to validate and import peptide matches into Integra database

Comments:
IP100013769 Mass: 49845 Total Score: 66 Peptides Matched: 2
Tax_id=9606 Alpha enolase, lung specific

Approve Match?	Query	Observed	Mr(exp)	Mr(Calc)	Delta	Miss	Score	Rank	Peptide
<input checked="" type="checkbox"/>	SZ	705.15	704.14	703.4	0.74	0	11	1	GVPLYR
<input checked="" type="checkbox"/>	346	714.38	1426.74	1424.72	2.02	0	56	1	Y69PQLADLYK

Proteins matching the same set of peptides:
IP100215736 Mass: 47481 Score: 66 Peptides Matched: 2
Tax_id=9606 enolase 1

Search Parameters

Done Local intranet

ABRF 2005



This filter will only display protein hits containing one or more peptide matches that include the consensus sequence for phosphorylation (serine or threonine followed by proline).

http://koala:8080/topaz/rc?command=page&page=ClusterSearchM5 - Microsoft Internet Explorer

Address http://koala:8080/topaz/rc?command=page&page=ClusterSearchM5#Cluster1

Cluster 1
Cluster Present in: mss-03022005-00001,mss-28012005-00001
Top Hit: q112876711 AL544231 LTI_NFL006_PL2 Homo sapiens cDNA clone CS00D1019YF12 S prime
Top Score: 953.414302532446

Cluster contains:

Accession	Mascot search id	Description	Protein Score	Mass	No peptides matched
q112814240 (frame 2)	mss-03022005-00001	AGENCOURT_10120424 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:6502033 5'	951.73	41667.55	19
q118526744 (frame 2)	mss-03022005-00001	AGENCOURT_6485545 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5556962 5'	722.86	40686.36	14
ANXI_PIG	mss-28012005-00001	Annexin A1 (Annexin 1) (Lipocortin 1) (Calpactin II) (Chromobindin 9) (F35) (Phospholipase A2 inh1b	61.64	38734.98	9
LUGP1	mss-28012005-00001	annexin I - guinea pig	41.24	38530.96	6
LUHU	mss-28012005-00001	annexin I - human	326.9	38689.98	42
SZ8Z8	mss-28012005-00001	annexin I - bovine	68.51	38873.22	9
ANXI_HUMAN	mss-28012005-00001	Annexin A1 (Annexin 1) (Lipocortin 1) (Calpactin II) (Chromobindin 9) (F35) (Phospholipase A2 inh1b	326.9	38558.94	42
q118512624 (frame 1)	mss-03022005-00001	AGENCOURT_6434022 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5518433 5'	861.14	38180.44	16
FAA64477	mss-28012005-00001	SSANNEXIN NID: - Sus scrofa	81.64	38215.77	9
Q75298	mss-28012005-00001	Hypothetical protein - Xenopus laevis (African clawed frog).	55.33	37612.45	9
q118512624 (frame 2)	mss-03022005-00001	AGENCOURT_6434022 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5518433 5'	861.14	38180.44	16
q112873366 (frame 2)	mss-03022005-00001	AL544874 LTI_FLD02_PL1 Homo sapiens cDNA clone CS0DE007YH11 S prime	950.05	37154.13	19
q112876711 (frame 2)	mss-03022005-00001	AL544231 LTI_NFL006_PL2 Homo sapiens cDNA clone CS00D1019YF12 S prime	953.41	37305.44	19
q112876711 (frame 1)	mss-03022005-00001	AL544231 LTI_NFL006_PL2 Homo sapiens cDNA clone CS00D1019YF12 S prime	953.41	37305.44	19
q112760114 (frame 1)	mss-03022005-00001	S023624307L NIH_MGC_30 Homo sapiens cDNA clone IMAGE:4471038 5'	468.06	33180.42	10
q113578471 (frame 3)	mss-03022005-00001	S0251945F1 NIH_MGC_79 Homo sapiens cDNA clone IMAGE:4719443 5'	237.31	25162.03	5
q113578471 (frame 1)	mss-03022005-00001	S0251945F1 NIH_MGC_79 Homo sapiens cDNA clone IMAGE:4719443 5'	237.31	25162.03	5
q110795811 (frame 3)	mss-03022005-00001	AV714294 DCB Homo sapiens cDNA clone DCBAAC07 5'	274.88	23469.71	7
q110795811 (frame 2)	mss-03022005-00001	AV714294 DCB Homo sapiens cDNA clone DCBAAC07 5'	274.88	23469.71	7
q110795811 (frame 1)	mss-03022005-00001	AV714294 DCB Homo sapiens cDNA clone DCBAAC07 5'	274.88	23469.71	7
q118622946 (frame 6)	mss-03022005-00001	QV1-HT0413-090200-062+06 HT0413 Homo sapiens cDNA	530.98	20864.82	11
q114392745 (frame 6)	mss-03022005-00001	HR2-HT1161-180101-002+01 HT1161 Homo sapiens cDNA	162.75	20241.36	5
q119881160 (frame 2)	mss-03022005-00001	K-EST0027758 S95NU601 Homo sapiens cDNA clone S95NU601-14-F22 5'	523.99	19285.02	10

Cluster 2
Cluster Present in: mss-03022005-00001,mss-28012005-00001
Top Hit: q113045775 AGENCOURT_10462366 NIH_MGC_109 Homo sapiens cDNA clone IMAGE:6445690 5'
Top Score: 482.20719116607216

Cluster contains:

Accession	Mascot search id	Description	Protein Score	Mass	No peptides matched
q113045775 (frame 2)	mss-03022005-00001	S02384694F1 NIH_MGC_39 Homo sapiens cDNA clone IMAGE:4513703 5'	389.6	39096.19	10
q111261496 (frame 2)	mss-03022005-00001	S02500048F2 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4129221 5'	182.98	36608.79	4

Another report uses BLAST to cluster protein hits from multiple searches into groups according to their peptide matches.

Custom reports

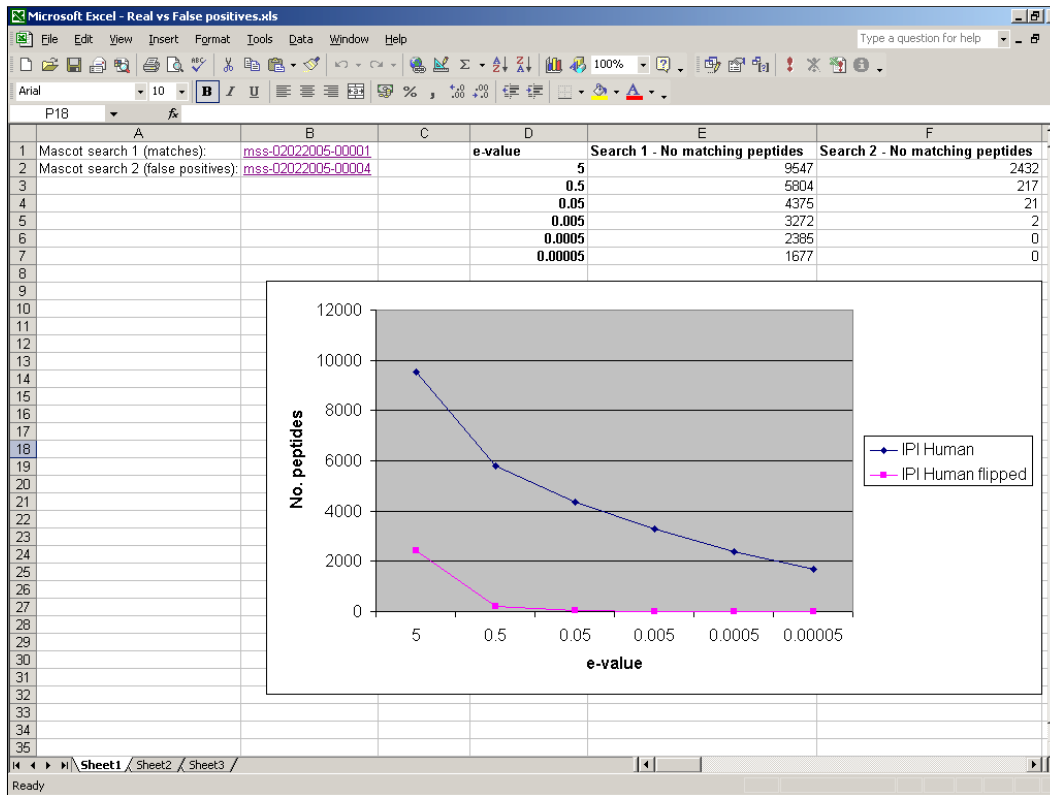
- Supports generation of Microsoft Excel reports
- Requires some knowledge of SQL query language
- Example reports will be supplied, users will be able to design and upload their own custom reports
- Database product – other reporting tools can be used

ABRF 2005

{MATRIX}
{SCIENCE}

The interface for custom reports is Microsoft Excel. Almost everyone has access to and some familiarity with Excel. It is a very powerful tool that allows you to put together and format all types of reports.

Mascot Integra comes with example reports and tutorials for building your own reports.



In this example of a custom report, a MudPIT dataset has been searched against the IPI Human database and also the same database with all the protein sequences reversed.

Any matches to the reverse database should be treated as false positives. Peptide matches are then extracted and plotted against their expectation values in Excel, displaying the true and false positive rates as a function of expectation value



Shipped as turn key system running on IBM xSeries 226

- Dual 3.2 GHz Xeon processors
- 2Gb RAM
- 4 x 146.8 GB SCSI RAID
- DDS/5 tape drive
- Windows Server 2003

Pricing based on

- number of named users

ABRF 2005



Mascot Integra ships as a turn-key system on IBM hardware

Summary

- Laboratory Information Management for proteomics
- *Not* an enterprise LIMS
- Oracle database
- All user functionality through a simple, clear web-browser interface
- Role and project/study membership based security
- Intuitive graphical experiment design
- Flexible custom reporting using Microsoft Excel.

ABRF 2005



Acknowledgements

- Mark Weeks, John Sinclair, Richard Jacob
 - Ludwig Institute for Cancer Research, London UK
- Katheryn Resing
 - University of Colorado, USA
- LabVantage Solutions Inc

ABRF 2005

