

*Mascot Integra:
Data management for
Proteomics*

ASMS 2004

**{MATRIX}
{SCIENCE}**

Mascot Integra: Data management for proteomics

- What is Mascot Integra?
- What Mascot Integra isn't
- Instrument integration in Mascot Integra
- Designing and running an experiment
- Results and reporting in Mascot Integra.

ASMS 2004



Mascot Integra is a new solution from Matrix Science for data and sample management in the proteomics laboratory.

What is Mascot Integra?

- Fully functional 'out of the box' solution for proteomics workflow and data management
- Support for all the major Mass-spec instrument manufacturers data systems
- Powered by the Sapphire® LIMS package from LabVantage Solutions inc
- Oracle 9 database
- Scaleable to the largest projects.

ASMS 2004

{MATRIX}
{SCIENCE}

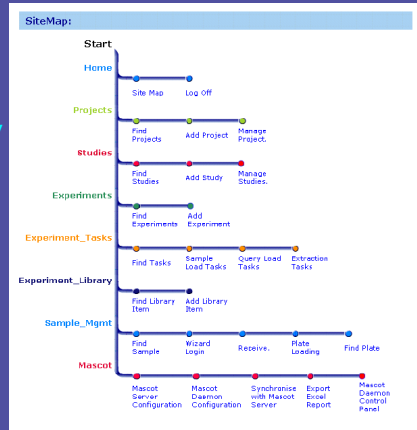
Mascot Integra is a fully functional application that will manage your proteomics workflow 'out of the box'. This contrasts with Enterprise LIMS applications, which require extensive customisation.

Rather than develop the LIMS functionality ourselves, we partnered with LabVantage Solutions Inc (www.lims.com). Their Sapphire LIMS package provides the core LIMS functionality for Mascot Integra

The Oracle database engine 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



ASMS 2004



Mascot Integra supports all aspects of the workflows associated with Proteomics laboratory processes.

In the initial release, NonLinear Dynamics Gel imaging software is supported

What Mascot Integra won't be

- **An enterprise LIMS**
 - Does not require a customised configuration after extensive consultancy
 - Won't run your entire organisation
 - Inventory management and reagent re-ordering etc
 - Won't run instrument data systems directly from the LIMS
- **As expensive as an enterprise LIMS.**

ASMS 2004

{MATRIX}
{SCIENCE}

Mascot Integra is intended to meet the needs of proteomics facilities and laboratories. It is not intended to be a substitute for a full scale LIMS system.

LabVantage Solutions Inc

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

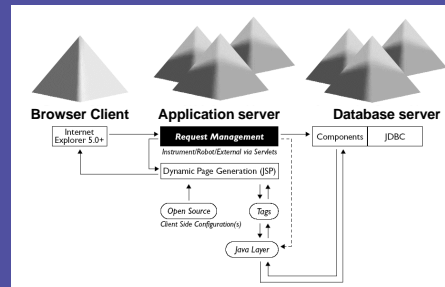
ASMS 2004

**{MATRIX}
{SCIENCE}**

If a Mascot Integra customer wishes to scale up to an Enterprise LIMS, or an existing Sapphire LIMS customer wishes to add Integra functionality, then LabVantage will be pleased provide the necessary services

Sapphire architecture

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



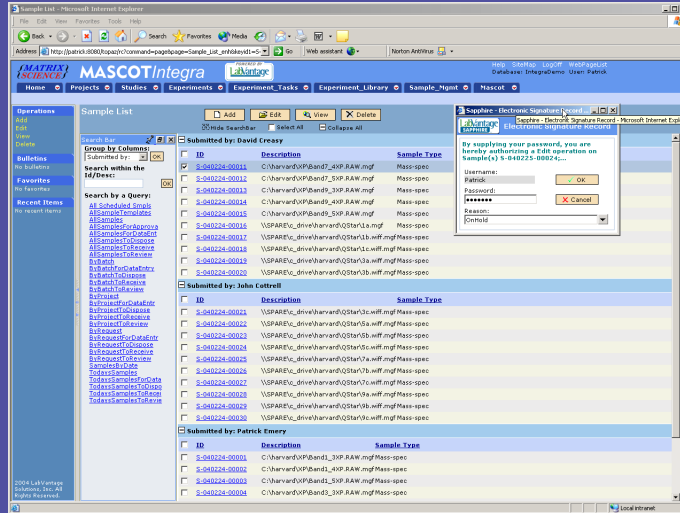
ASMS 2004

**MATRIX
SCIENCE**

Sapphire has a true 3 tier architecture. The client interface is 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)



ASMS 2004



Mascot Integra meets the requirements of CFR21 part 11 ERES

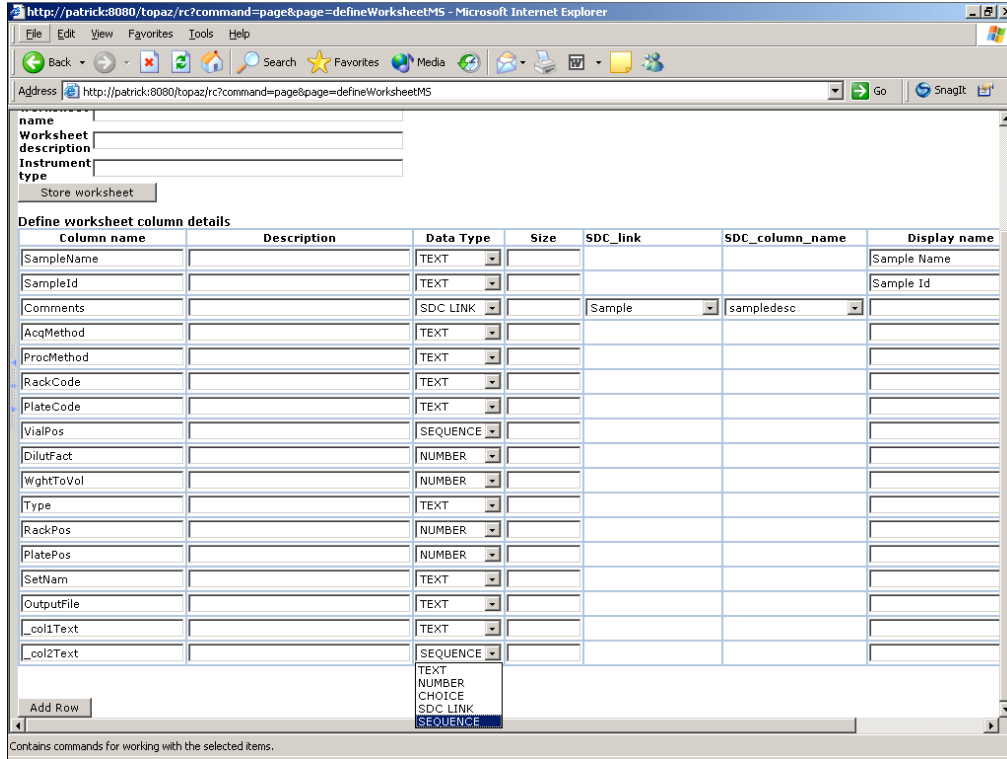
Instrument Integration

- Not running the instrument from the LIMS
- Handled via Sample / work sheet exchange
- Comes with a range of worksheets for the main instrument types, data systems and manufacturers
- Flexible – you will be able to design your own worksheets and templates
- Output CSV, TSV and Microsoft Excel files.

ASMS 2004

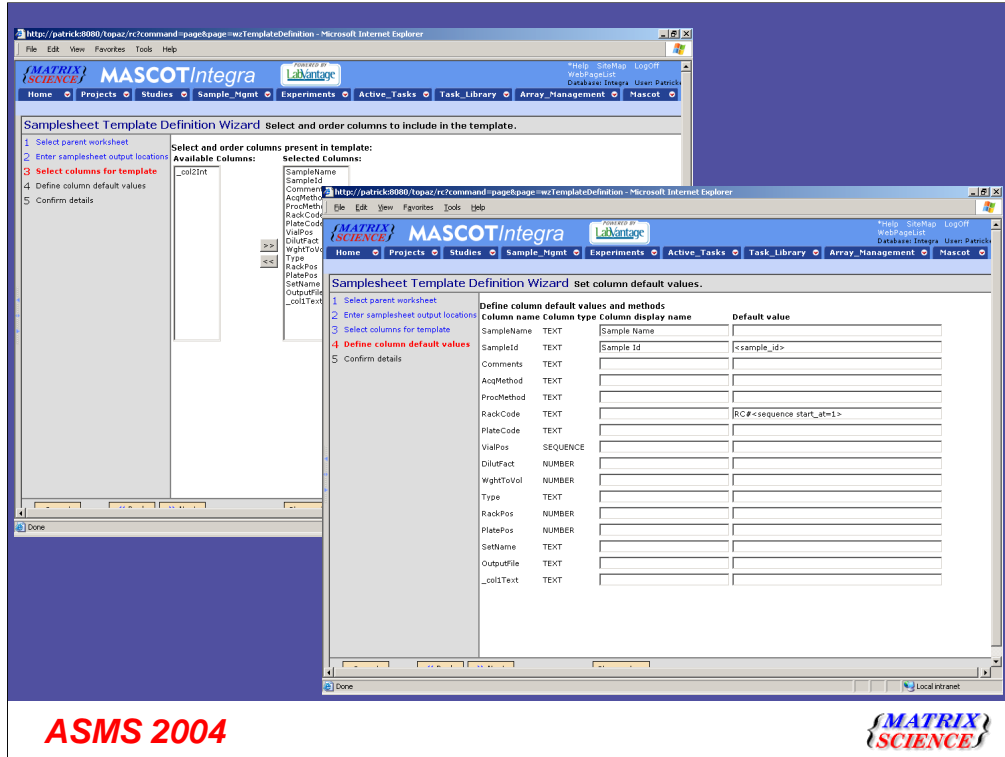
**{MATRIX}
{SCIENCE}**

Integration with complex instruments, such as mass spectrometers, is handled via sample sheet exchange. It is not desirable to attempt low level control of systems that have their own data system.



To design a new sample sheet, you create a parent worksheet specifying the required columns.

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 link back to data held in the LIMS database tables. The data retrieved can be limited by the sample identifier associated with the row of the sample sheet being created.



Once the parent worksheet is defined, you design a template (or view) for the sample list itself.

You select the columns required for the template and can change their display order as required. You can also specify default values for the columns

Running a basic experiment:

- Now we'll go through setting up and running a basic experiment:
 - Loading a 96 well plate
 - Setting up a Mass-Lynx sample sheet
 - Defining raw-data processing with Mascot Distiller
 - Setting up a Batch search with Mascot Daemon+
 - Running the Mascot searches
 - Parsing the results into the database
 - Examining the results

ASMS 2004

**{MATRIX}
{SCIENCE}**

I'll describe the steps involved in running a basic experiment. Although there may look to be a large number of steps, it is quick and easy in practice

Experiment design:

- Experiments are broken into a series of experimental tasks
- Experimental tasks have inputs and outputs (e.g. 96 Well 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
- A comprehensive library of flexible proteomics tasks is provided.

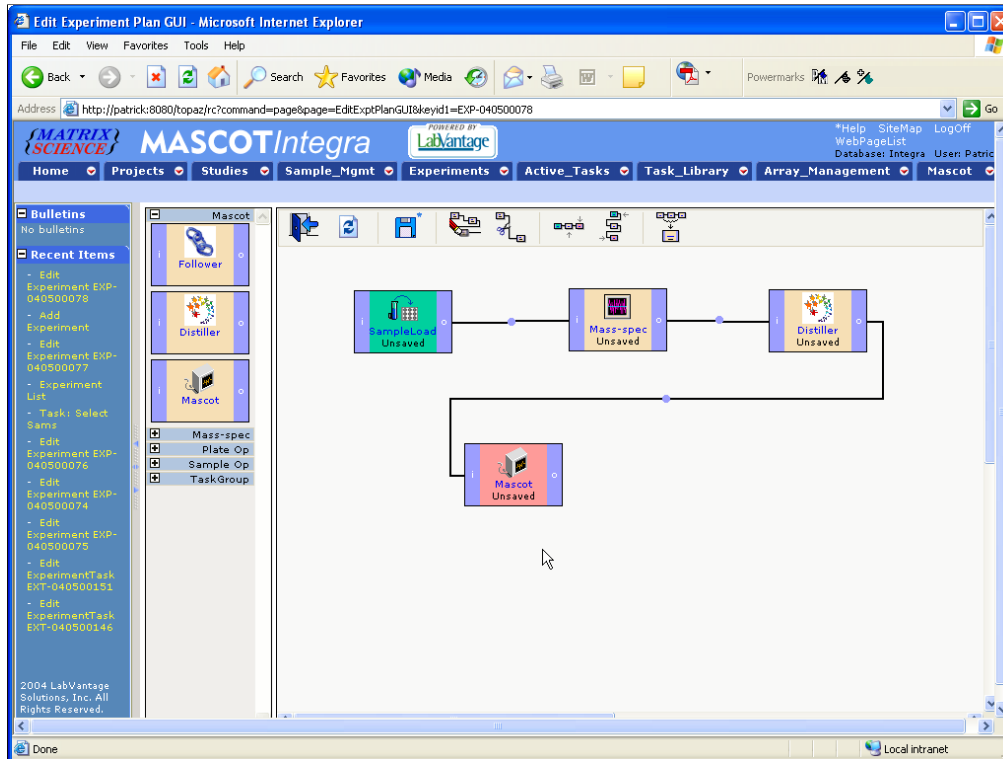
ASMS 2004

{MATRIX}
{SCIENCE}

Experimental tasks represent laboratory processes e.g. running a 2D gel, loading a 96 well plate, sample digestion etc

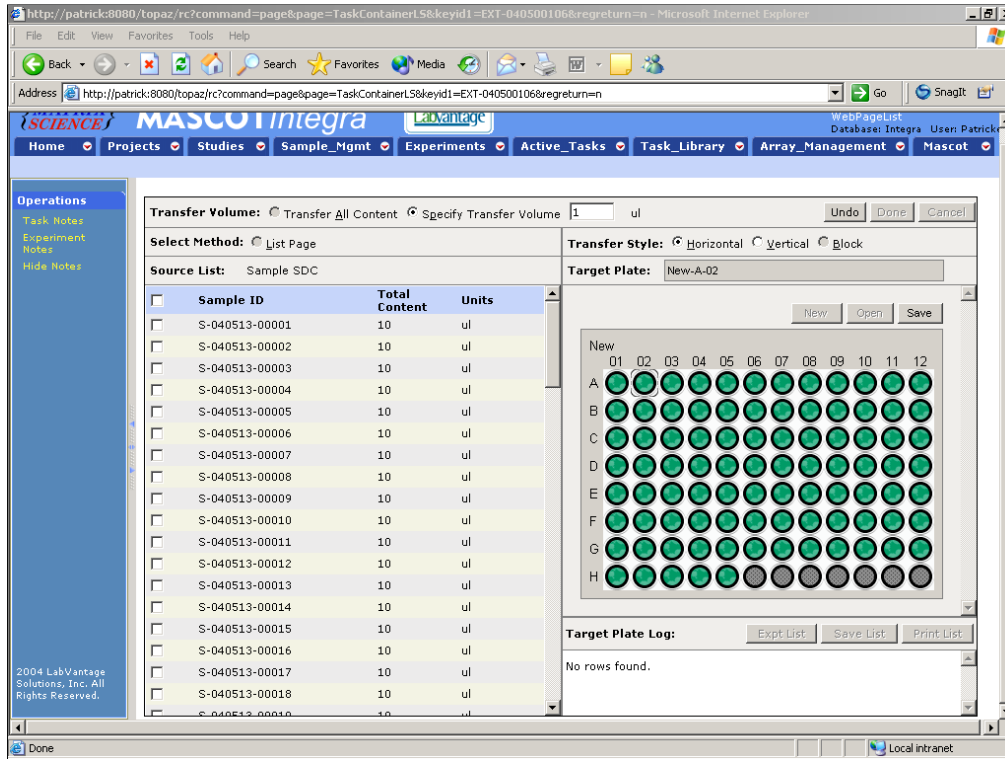
For an experimental task of loading a 96 well MALDI plate, the inputs would be a 96 well MALDI plate and a sample list, and the output would be a loaded 96 well MALDI plate.

Experimental tasks can have multiple inputs and outputs.



To define a workflow, you simply drag and drop the required icons onto the desktop then connect them using the mouse.

To save time in the future, each new workflow can be saved as a new experimental task. These tasks are displayed in the task list to the left of the main design area. Choosing a task automatically re-creates the pre-defined workflow.



The first step is to load the samples from a 96 well plate. Flexible support is provided for two dimensional sample holders, such as microtitre plates and mass spectrometry sample targets

Task 2: Defining a MassLynx 4.0 sample sheet

Select a pre-defined template based on a MassLynx 4.0 samplesheet

Some data is automatically filled in from the database

Export as CSV or xls format file for import into MassLynx

Cell No.	Index	Version	File Name	File Text	MS File	MS Toss File	Inlet File	Inlet C
A-01	1		S-040504-00007					
A-02	2		S-040504-00008					
A-03	3		S-040504-00009					
A-04	4		S-040504-00100					
A-05	5		S-040504-00101					
A-06	6		S-040504-00102					
A-07	7		S-040504-00103					
A-08	8		S-040504-00104					
A-09	9		S-040504-00105					
A-10	10		S-040504-00106					
A-11	11		S-040504-00107					
A-12	12		S-040504-00108					
B-01	13		S-040504-00109					
B-02	14		S-040504-00110					
B-03	15		S-040504-00111					
B-04	16		S-040504-00112					
B-05	17		S-040504-00113					

ASMS 2004

MATRIX SCIENCE

The second step is to define the sample list. This will define how the samples are analysed in the mass spectrometer

Task 3: Peak detection with Mascot Distiller

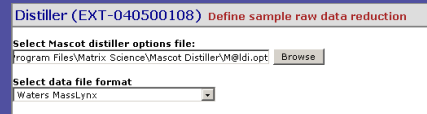
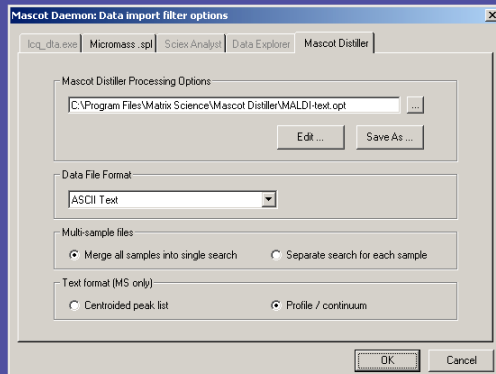
- Mascot distiller used for automated peak detection.
- Used in conjunction with Mascot Daemon for automated search submission.
- Peak detection via other software is supported.

ASMS 2004



Once the samples have been analysed, the third step is to process the MS data files into optimised peak lists using Mascot Distiller.

Peak detection with Mascot distiller continued...



ASMS 2004



Peak detection settings are pre-defined.

Task 4: Mascot search submission

- Automated search submission using Mascot Daemon
- Multiple Mascot daemon servers running as services
- A single TaskDB instance on the central Oracle server
- Parameter sets are held in the central TaskDB, not in local parameter files

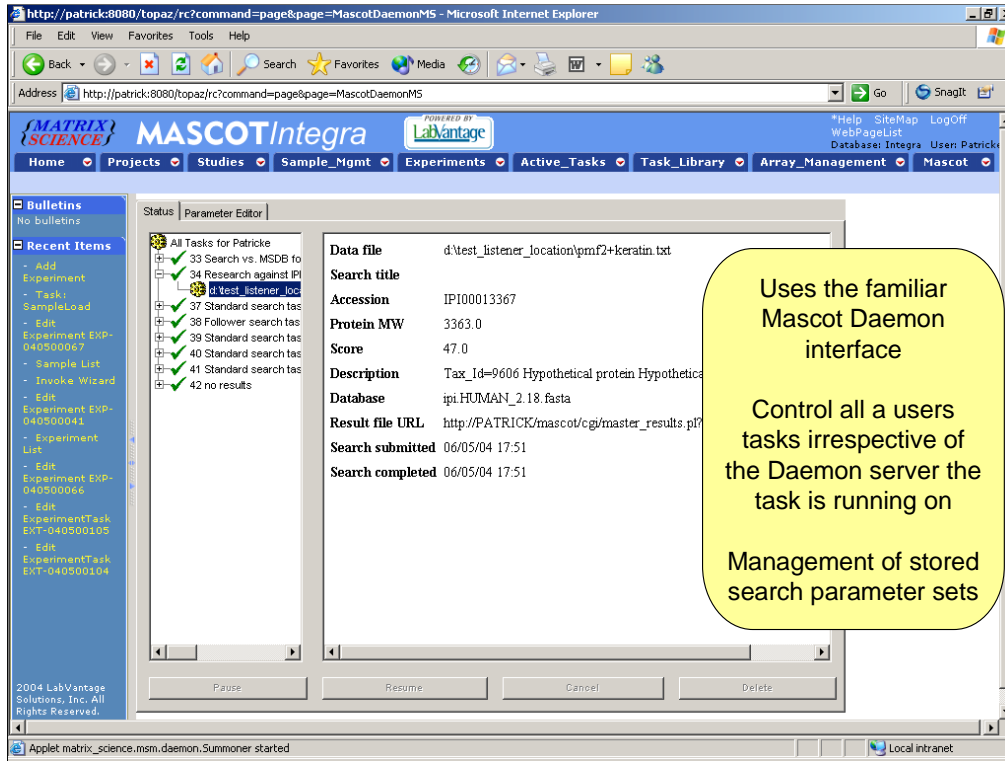
ASMS 2004



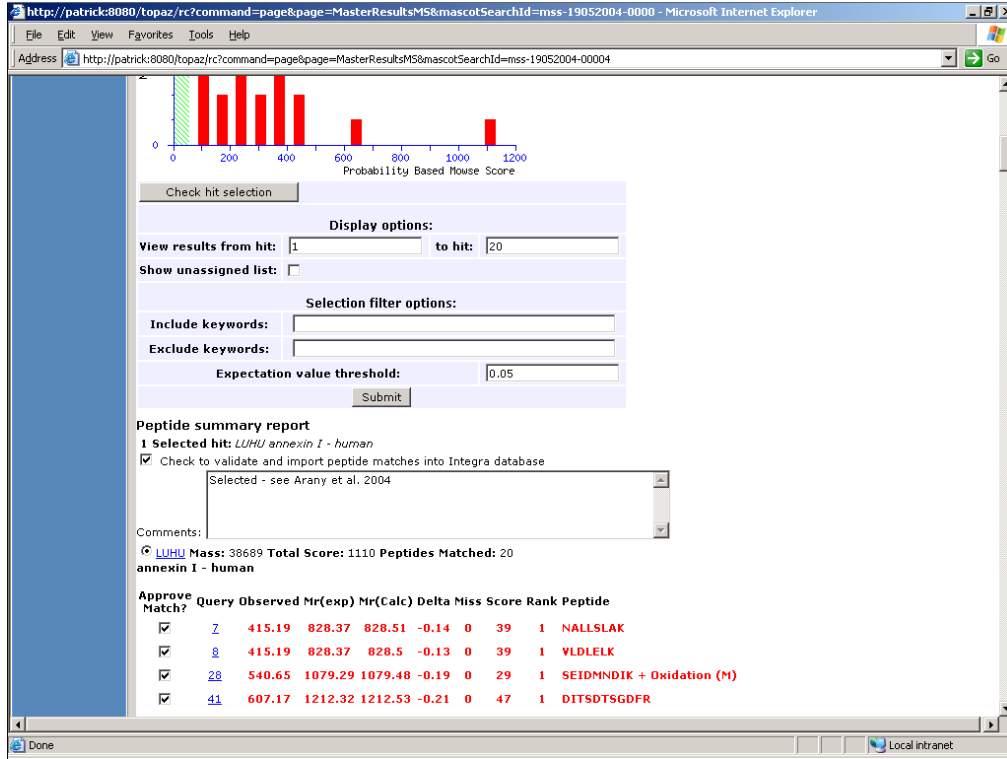
Mascot search submission is accomplished by a specialised version of the Mascot Daemon.

Each instrument data system has its own Mascot Daemon client. These all communicate with tables in the central LIMS database. Each Daemon client can be controlled from the Mascot Integra browser interface.

Search parameter sets are held centrally. This ensures that the parameters used by all the Mascot Daemon clients are identical.



A Java Applet that simulates the familiar Mascot Daemon interface is used for monitoring and controlling searches, irrespective of which Mascot Daemon client the tasks are running on, and for managing the centrally stored parameter sets.



Once a Mascot search is complete, the results are automatically parsed into the database tables. It is possible to generate all the standard Mascot html reports direct from these tables.

In addition, it is possible to annotate the reports. For example, approve a particular protein or peptide match or add comments.

http://patrick:8080/topaz/rc?command=page&page=ProteinViewMS&mascotSearchId=mss-19052004-00004& - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address http://patrick:8080/topaz/rc?command=page&page=ProteinViewMS&mascotSearchId=mss-19052004-00004&accession=LUHU&showAll=false

List
 - Add
 - Experiment
 - Task: Select
 - Sams
 - Edit
 - Experiment EXP-040500077
 - Edit
 - Experiment EXP-040500076

2004 LabVantage Solutions, Inc. All Rights Reserved.

Nominal mass (Mr): **38689.98**; Calculated pI value: **6.57**

LUHU has previously been selected by:
 Patricke on 20-May-2004 12:27:41. Selected - see Arany et al. 2004

NCBI BLAST search of **LUHU** against nr
 Unformatted [sequence string](#) for pasting into other applications
 Variable modifications: Carbamidomethyl (C),Oxidation (M),Propionamide (C),Pyro-glu (N-term E),Pyro-glu (N-term Q)

Cleavage by Trypsin/P: cuts C-term side of KR
 Total number of queries searched: 166
 Total number of queries matched: 20
 Sequence Coverage: 51.4%

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

1 MAMVSEFLKQ **AWFIENEEOE YVQTVKSSKG GPGSAVSPYP TFPSSDVAA**
 51 **LHKAIHVKGV DEATIIDILT KRNNQRQOI KAAYLQETGK PLDETLLKAL**
 101 TGHLEVVLA LLK**TPAQFDA DELRAAMKGL GTDEDTLIEI LASR**TNKEIR
 151 DINRVYREEL **KRDLAQDITS DTSGDPRNAL I.SLAKGDRSE DFGVNEIDLAD**
 201 **SDARALYEAG ERRKGTQVNV FNTLITRYS** PQLRRVFRQY TKYSKHDHNR
 251 **VLDLELKGDI EKCLTAIVKC ATSKPAFFAE** KLHQAMKGVG TRHKALIRIM
 301 **VSRSEIDHND IKAFYQRMYG ISLCQAILDE** TKGDYERILV ALCOGN

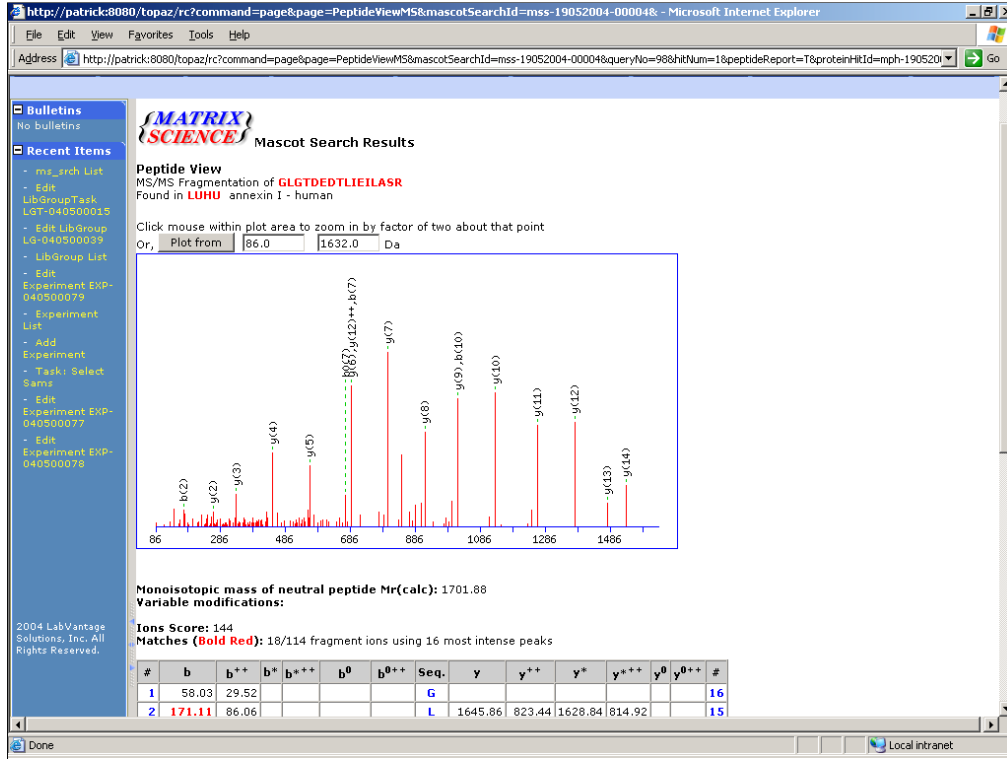
Show predicted peptides

Sort Peptides By Residue Number Increasing Mass Decreasing Mass

Start	End	Observed	Mr(exp)	Mr(calc)	Delta	Miss	Sequence	Ions score
10	26	1070.82	2139.62	2140.01	-0.39	0	QAWFIENEEOEYVQTVK	(92.65)
30	53	785.92	2354.73	2355.15	-0.42	0	GGPGSAVSPYPPTFPSSDVAA LHK	(110.99)
59	71	694.26	1386.5	1386.76	-0.26	0	GVDEATIIDILTK	(82.28)
59	72	515.2	1542.58	1542.86	-0.28	1	GVDEATIIDILTKR	(63.46)
82	98	476.93	1903.67	1904.03	-0.35	2	AAYLQETGKPLDET LKK	(41.94)
114	124	631.7	1261.38	1261.59	-0.21	0	TPAQF DA DELRL	(89.83)
125	144	707.55	2119.64	2119.08	0.56	1	AANKGLGTDEDTLIEILASR + Oxidation (M)	(28.82)
129	144	851.77	1701.53	1701.88	-0.35	0	GLGTDEDTLIEILASR	(144.04)
163	177	820.75	1639.48	1639.77	-0.28	1	DLAKDITS D TSGD F R	(63.8)
167	177	607.17	1212.32	1212.53	-0.21	0	DITS D TSGD F R	(47.63)

Done Local intranet

Previously added comments are flagged up on the protein view page.
 The blue background for the peptide matches shows that these peptide matches were all selected.



A Peptide View report generated from tables, not from the Mascot result file

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 (e.g. Business Objects) can be used.

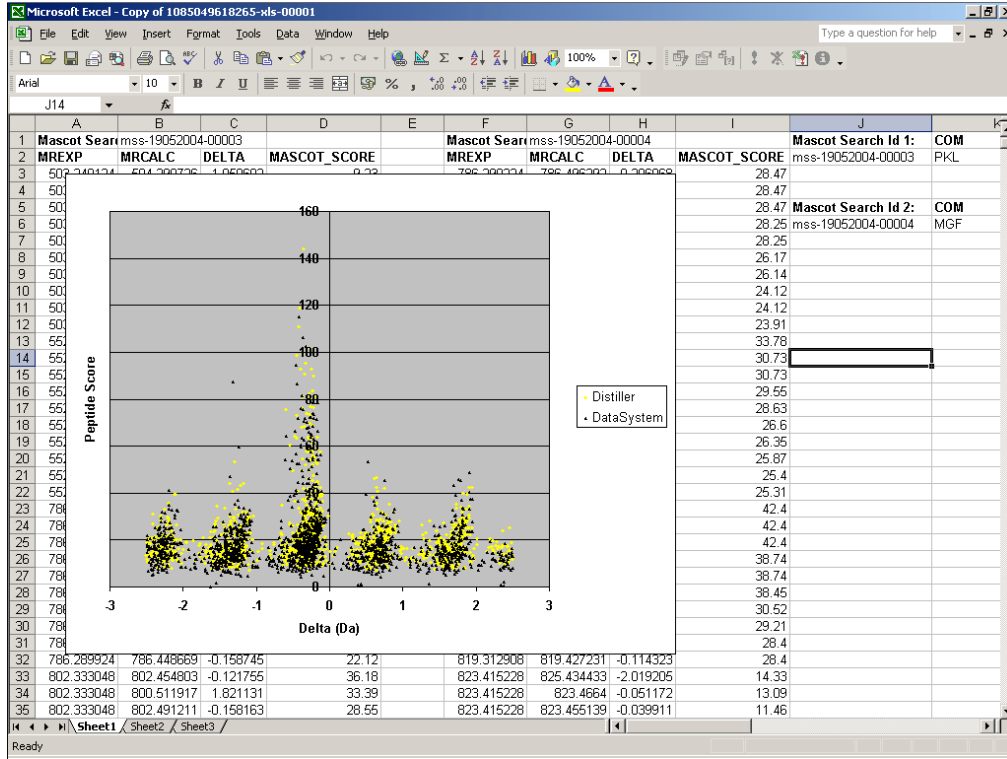
ASMS 2004

{MATRIX}
{SCIENCE}

With the Mascot search results in the database, reporting options are greatly expanded. We have chosen Microsoft Excel as the reporting tool / report designer.

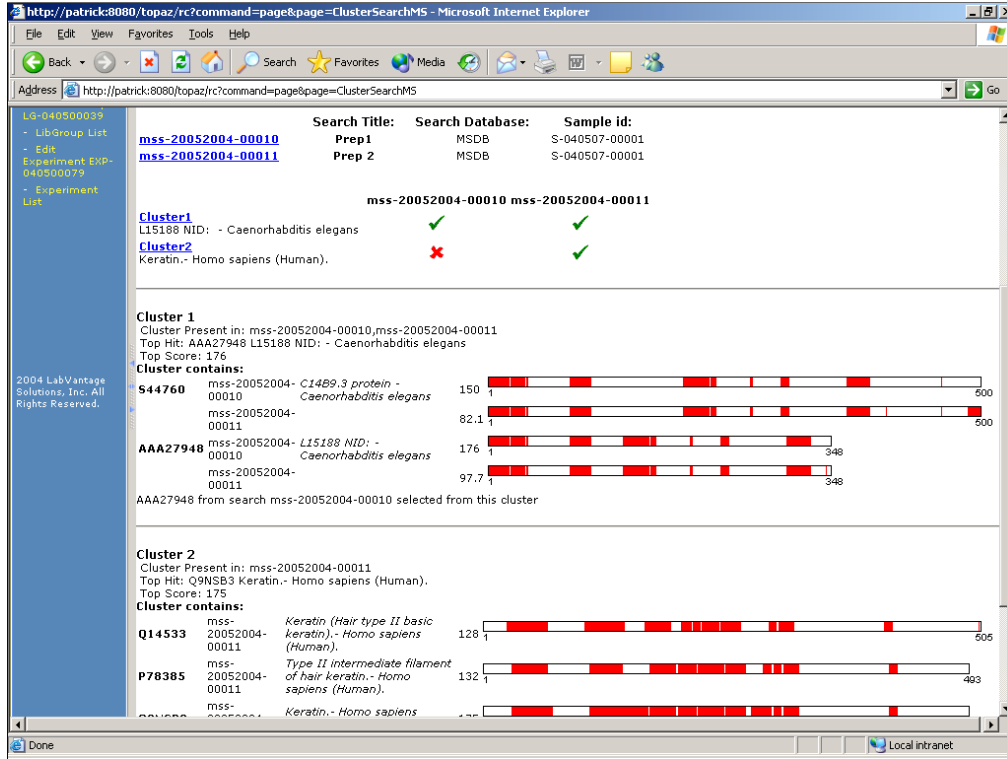
Almost everyone is familiar with Excel, and it has very powerful reporting capabilities.

If you have standardised on some other reporting tool, you will still be able use it. However, Mascot Integra will only include templates and documentation for Excel



As a simple example, this report plots mass error against score for two different peak lists.

Excel makes it easy to add features such as graphs. Mascot Integra ships with a number of templates to illustrate how tabular and graphical reports can be created.



Another report that is available through the web application enables protein hits from multiple Mascot searches to be clustered together to produce a summary report. The report highlights the locations of peptides common to each set of proteins.

Summary

- Laboratory Information Management for proteomics
- *Not* an enterprise LIMS
- Oracle database
- All user functionality through a simple, clear web-browser interface
- Intuitive graphical experiment design
- Flexible custom reporting using Microsoft Excel
- Target release date late 2004.

ASMS 2004

MATRIX
SCIENCE