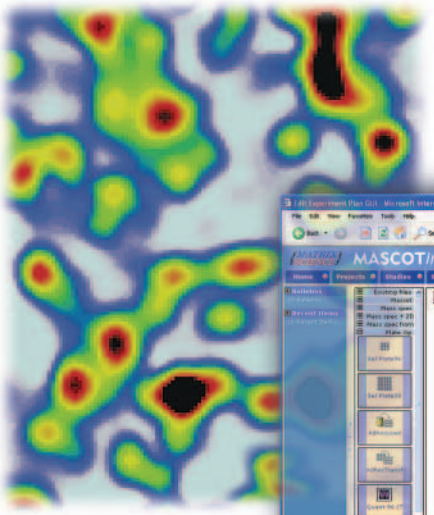
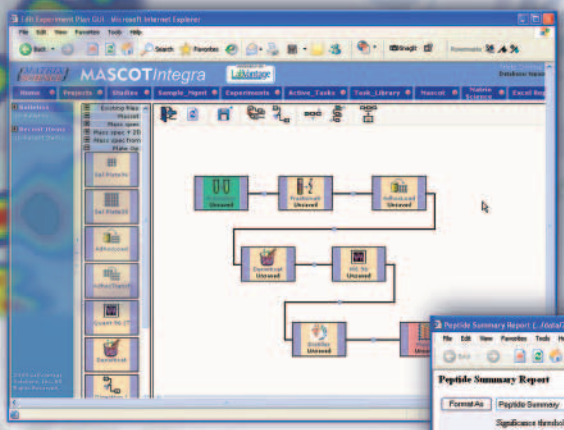
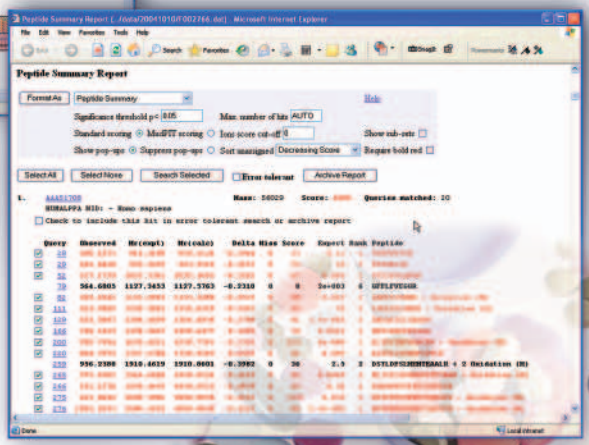
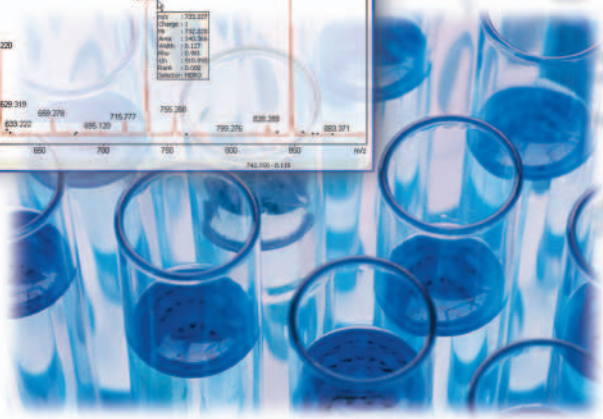
MASCOT®

Take the guesswork out of protein identification...









{MATRIX SCIENCE}

www.matrixscience.com

MASCOT®: Software for Protein Identification Using Mass Spectrometry Data

Proteomics is the study of the protein complement of a cell or tissue in a specific physiological state. The core technologies of proteomics are gel electrophoresis and HPLC for separation followed by mass spectrometry for analysis. Mascot takes the mass spectrometry data and searches it against molecular sequence databases to identify the constituent proteins and to characterize post-translational modifications. The search procedure is computationally intensive, requiring complex statistical calculations to be performed rapidly while streaming through protein or nucleic acid sequence databases.

Mascot has become established as the *de facto* standard for this application, because of advantages such as these:

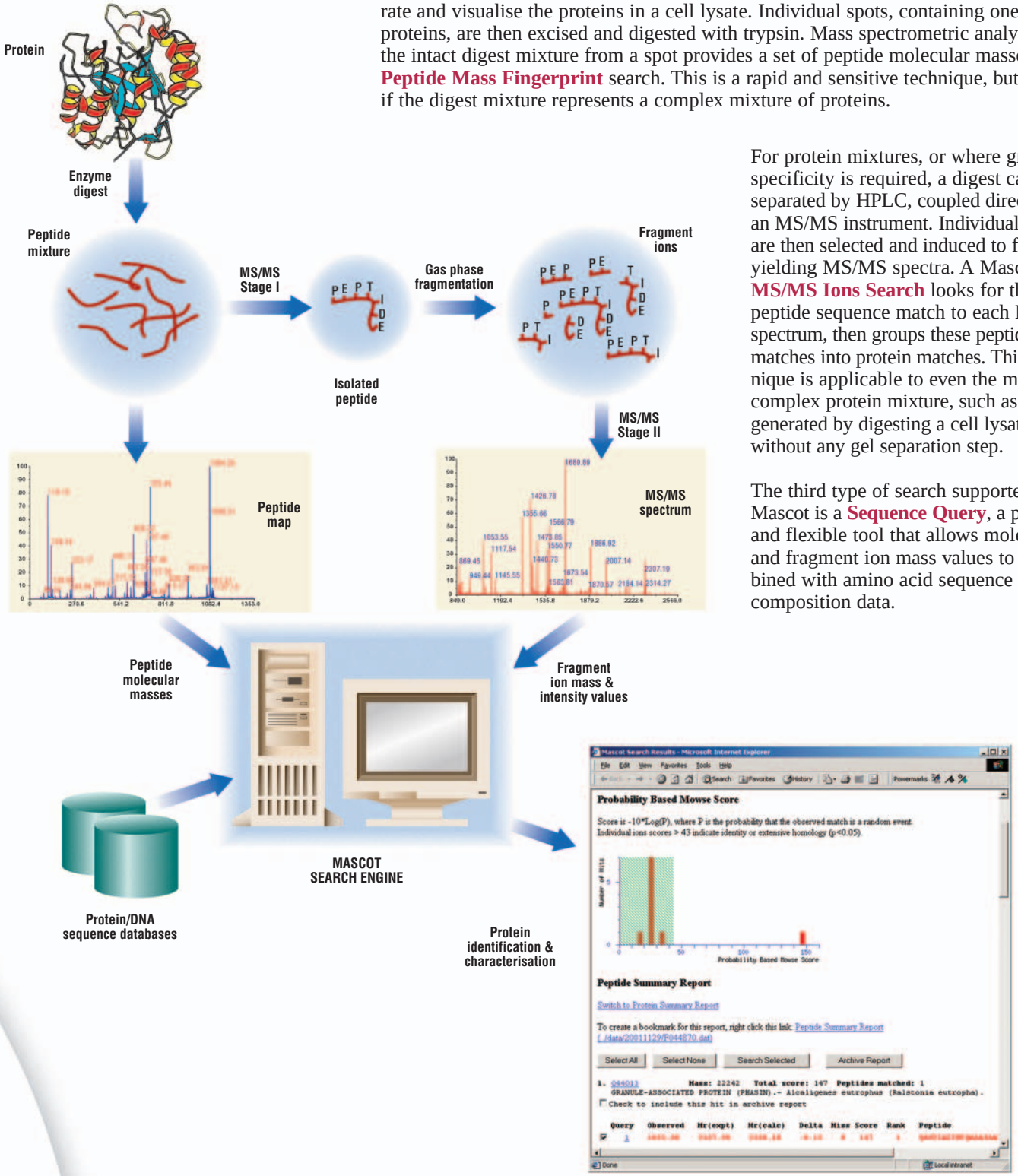
- Supports all three proven search strategies in a single, integrated package
- Unique, true probability based scoring allows standard statistical tests of significance to be applied.
- Search any FASTA database, whether protein, EST, or genomic DNA.
- No time-consuming index building; total flexibility in specifying chemical or post-translational modifications; search with or without enzyme specificity.
- Accepts mass spectrometry data from all the leading instrument manufacturers
- Fast, threaded code gives high throughput on a wide range of single and multi-processor systems and clusters.
- Sophisticated client software to automate search submission without custom programming.
- Summary and detail reporting of search results to any web browser, together with comprehensive, on-line help.
- Licensed for in-house use by more than a thousand academic and commercial laboratories. Described by Frost & Sullivan as “the gold standard for the searching of databases with mass spectrometric data”.
- Supported by a dynamic and independent company, committed to developing state-of-the-art bioinformatics software.

Supports All Three Proven Search Strategies in a Single, Integrated Package

One of many approaches to proteome analysis is to use 2D gel electrophoresis to separate and visualise the proteins in a cell lysate. Individual spots, containing one or a few proteins, are then excised and digested with trypsin. Mass spectrometric analysis of the intact digest mixture from a spot provides a set of peptide molecular masses for a **Peptide Mass Fingerprint** search. This is a rapid and sensitive technique, but may fail if the digest mixture represents a complex mixture of proteins.

For protein mixtures, or where greater specificity is required, a digest can be separated by HPLC, coupled directly to an MS/MS instrument. Individual peptides are then selected and induced to fragment, yielding MS/MS spectra. A Mascot **MS/MS Ions Search** looks for the best peptide sequence match to each MS/MS spectrum, then groups these peptide matches into protein matches. This technique is applicable to even the most complex protein mixture, such as those generated by digesting a cell lysate without any gel separation step.

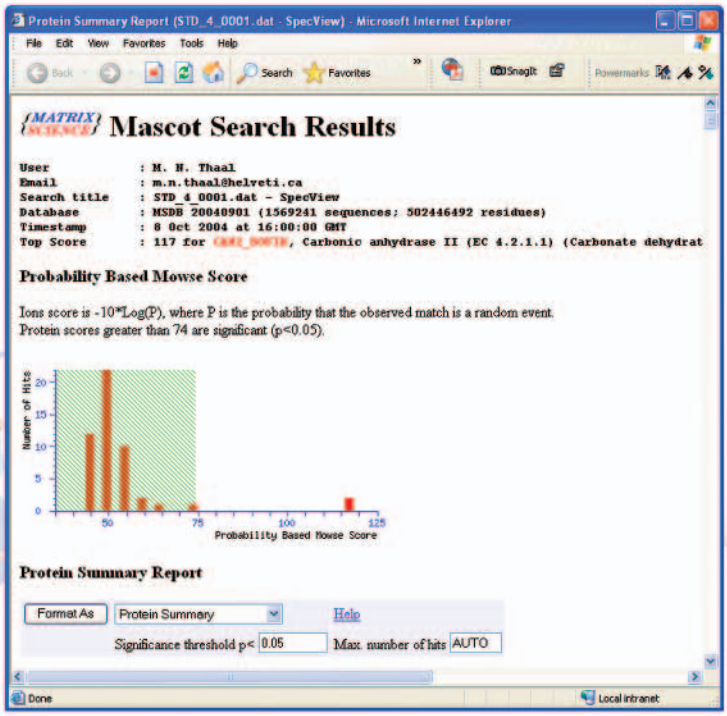
The third type of search supported by Mascot is a **Sequence Query**, a powerful and flexible tool that allows molecular and fragment ion mass values to be combined with amino acid sequence and composition data.



Unique, True Probability Based Scoring Allows Standard Statistical Tests of Significance to be Applied

Mascot computes the probability that the observed match between the experimental data and mass values calculated from a candidate peptide or protein sequence is a random event. The correct match, which is not a random event, then has a very low probability.

True probability based scoring is the key to recognising and avoiding false positives. It is also an essential pre-requisite for automation. Only by establishing scores on a fixed, absolute scale, can the decision to accept or reject an identification be made by simple, rule-based software.



A histogram of the Mascot score distribution for the top 50 best matching proteins is displayed at the top of a peptide mass fingerprint report. Scores are -10*Log(P), where P is the probability that the observed match is a random event. Scores in the green shaded area represent random matches, while the correct match has a score of 117. The chance of this being a random match is 3 in 10^6.

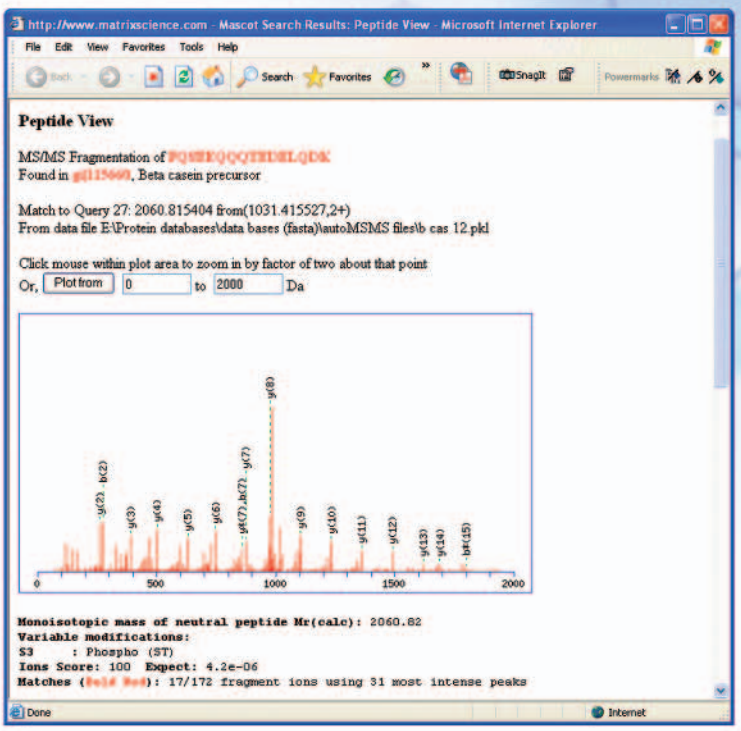
Reference: Perkins, D.N. et al. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, 20, 3551-3567

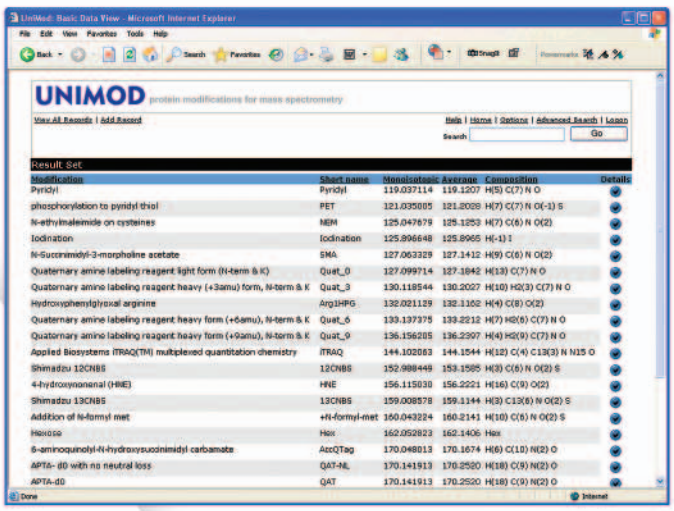
Search Any FASTA Database, Whether Protein, EST, or Genomic DNA

All calculations are performed on the fly, direct from the FASTA file, giving total flexibility.

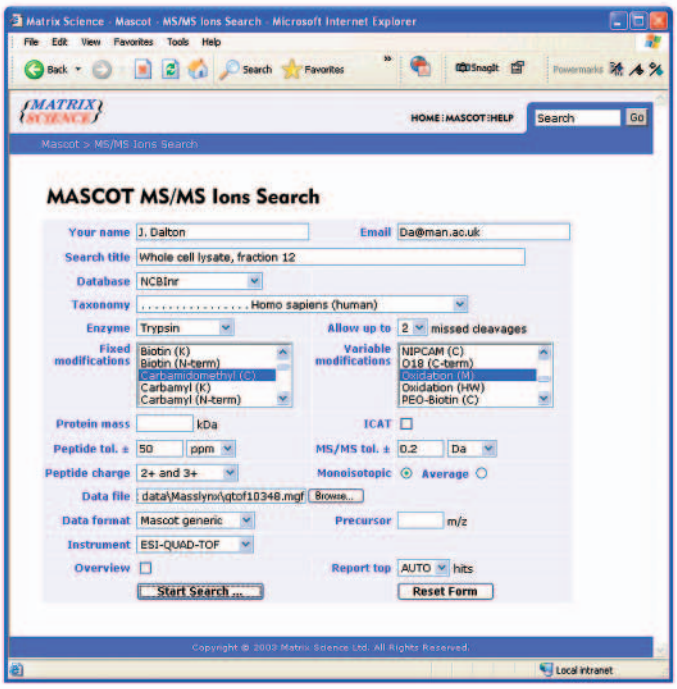
Search with or without enzyme specificity. 'No-enzyme' searches using MS/MS data are essential for finding non-specific cleavage products, and working with targets such as MHC peptides.

A wide range of chemical and post-translational modifications can be specified in a search. These modifications can be fixed (quantitative) or variable (non-quantitative). Arbitrary combinations of fixed modifications and up to 9 variable modifications can be included in a single search. If a peptide contains multiple potential modification sites, as in this example, Mascot can identify precisely which residues have been modified.





The choice of modifications is taken from a simple text file, which can be updated by editing or by downloading a new file from www.unimod.org. New modifications are available for searching immediately, no waiting for new indexes to be built.



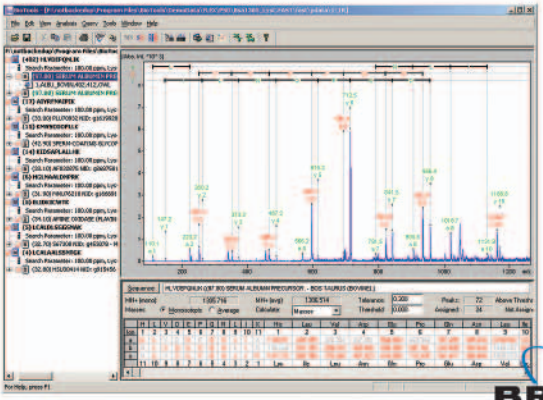
For interactive searching, the user interface is provided by any JavaScript aware web browser.

Mascot accepts Mass Spectrometry Data from all the Leading Instrument manufacturers

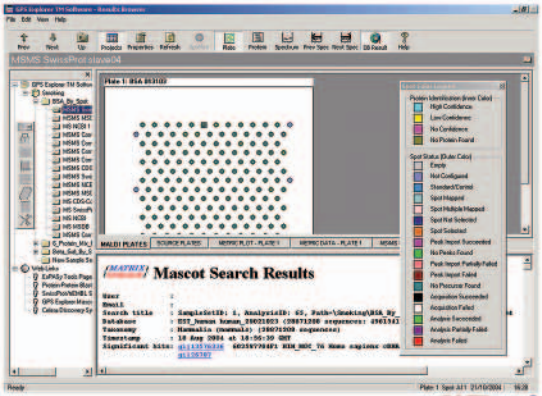
Protein and peptide analysis by mass spectrometry uses a bewildering variety of instruments and techniques, which means that spectra can vary enormously in terms of mass resolution, signal to noise, and fragmentation behaviour.

Mascot isn't tuned for data from one particular type of instrument. It is designed to extract all the statistically significant information and deliver optimum results, whatever the source of the data.

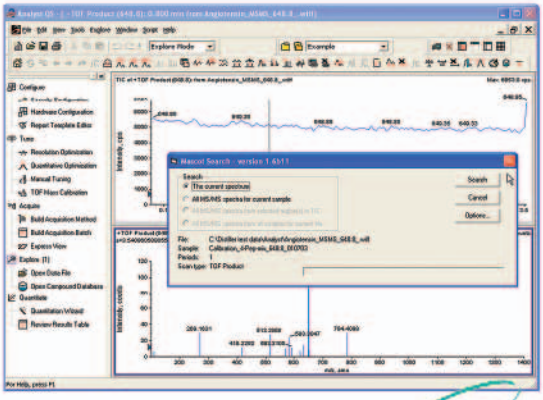
Most instrument vendors now incorporate a Mascot interface into their data analysis packages. In addition, you can always use Mascot Daemon to automate data processing and search submission, as illustrated elsewhere in this brochure for Finnigan Xcalibur data from Thermo ELECTRON CORPORATION



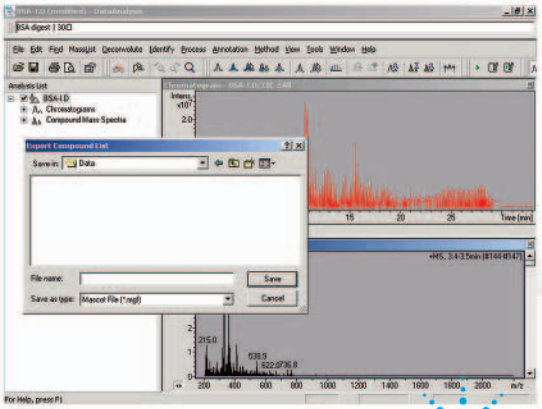
BRUKER



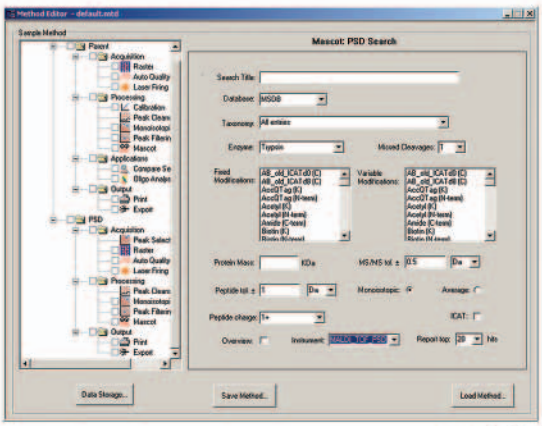
Applied Biosystems



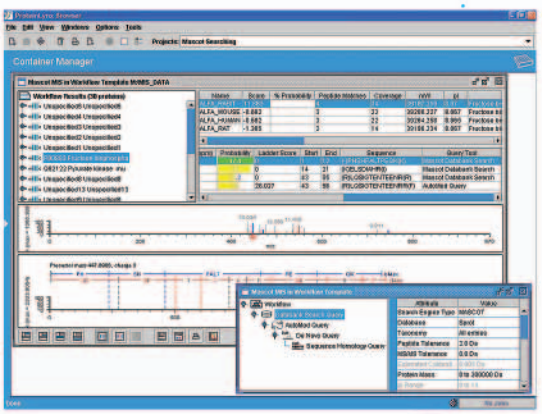
MDS SCIEX



Agilent



KRATOS ANALYTICAL A SHIMADZU GROUP COMPANY

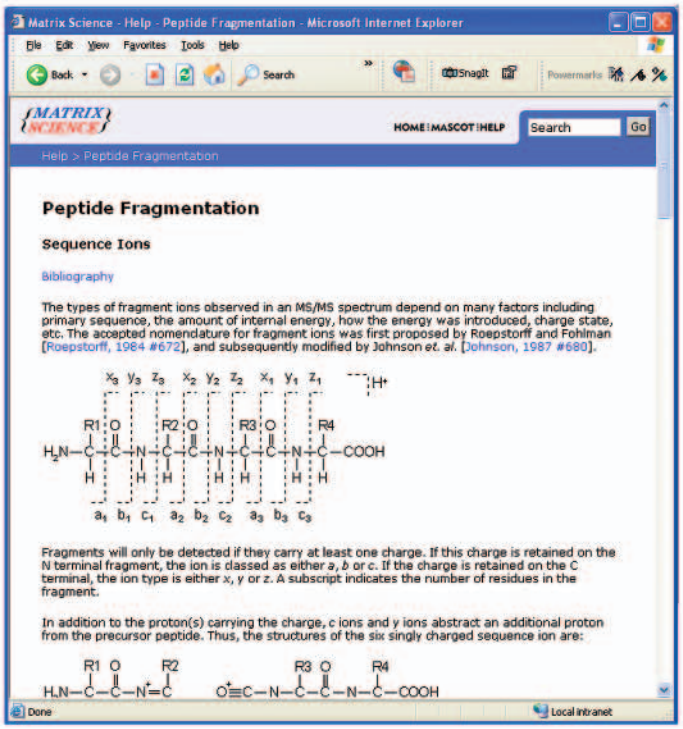


MICROMASS

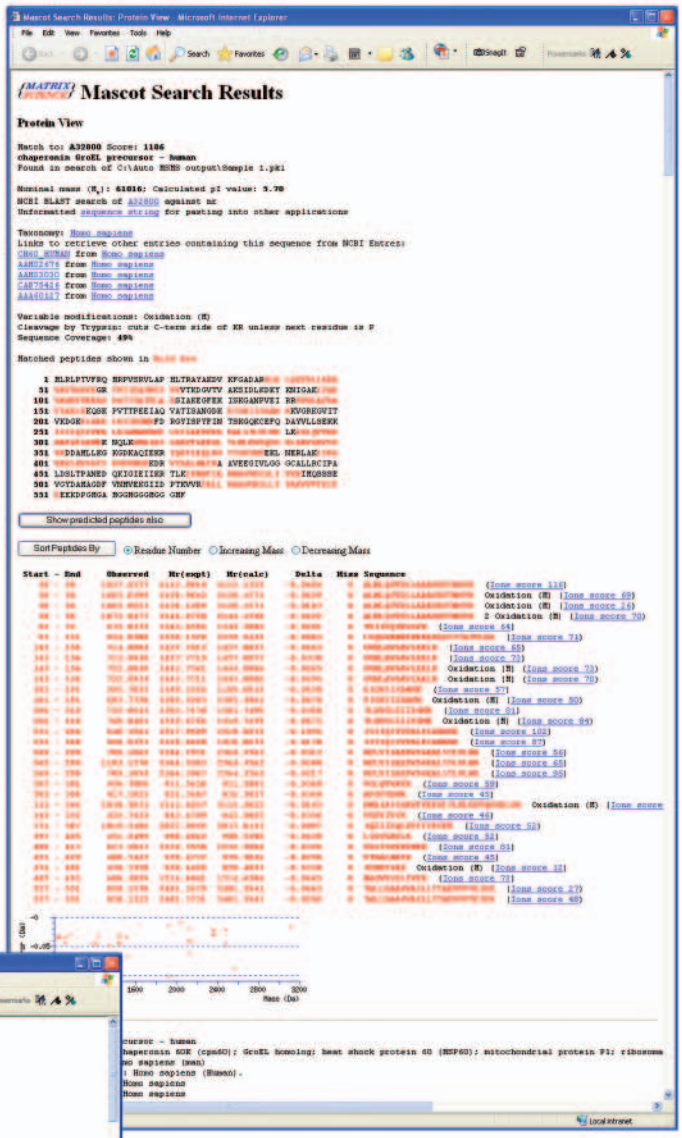
All trade marks and service marks on this page are the properties of their respective owners and are hereby acknowledged

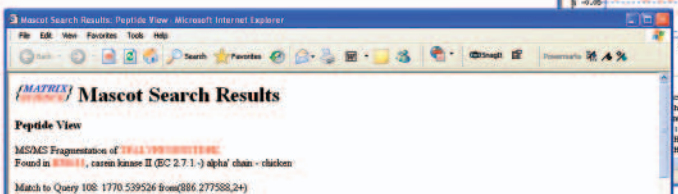
Summary and Detail Reporting of Search Results to any Web Browser, Together With Comprehensive, On-line Help

Comprehensive reports help visualise and understand the search results. The example to the right shows a protein view report from a search of LC-MS/MS data.



Detailed on-line help is provided in the form of HTML pages





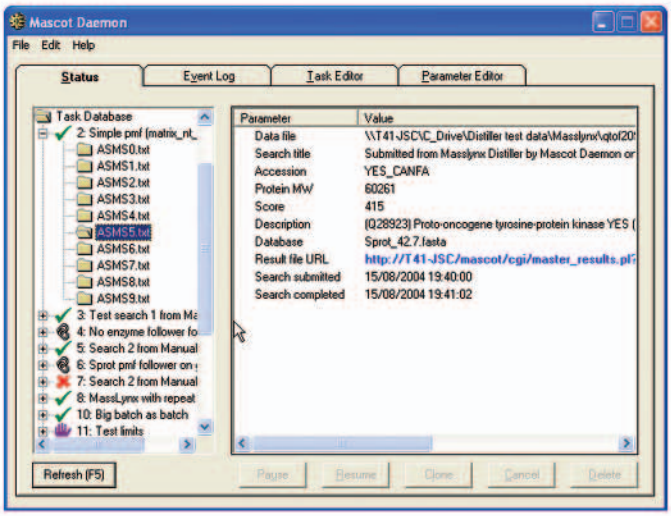
By clicking on a link, you can drill down details of an individual peptide match

Sophisticated Client Software to Automate Search Submission Without Custom Programming

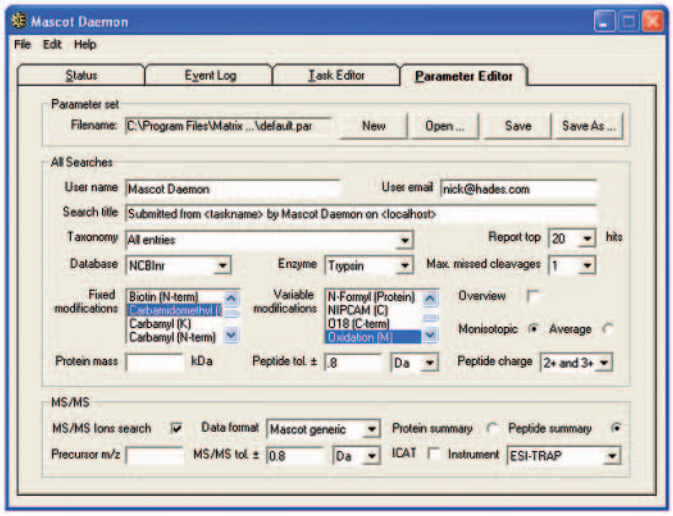
Data streams from multiple mass spectrometers can be routed to a Mascot Server for real-time searching by using Mascot Daemon, a Microsoft Windows application that is bundled with Mascot.

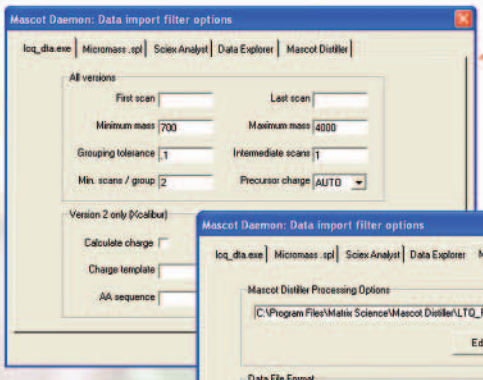
Each Mascot Daemon task defines the data source (a list of data files or a file path), how the data are to be searched, when the searches are to take place, and any follow-up activities, such as conditional repeat searches.

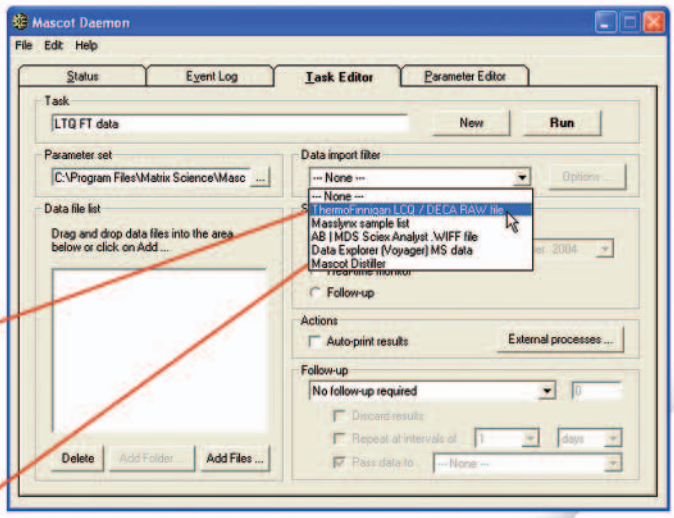
- Batch task: A batch of data files to be searched immediately or at a defined time
- Real-time monitor task: New files on a defined path are searched as they are created.
- Score dependent follow-up task. For example, automatically repeating a search against a different sequence database.



Search parameters are defined in the Parameter Editor, which closely resembles the HTML form used for interactive Mascot searches. Tasks and their associated search results are displayed on the status tree. The full result report can be displayed in a web browser by clicking on the blue hyperlink.







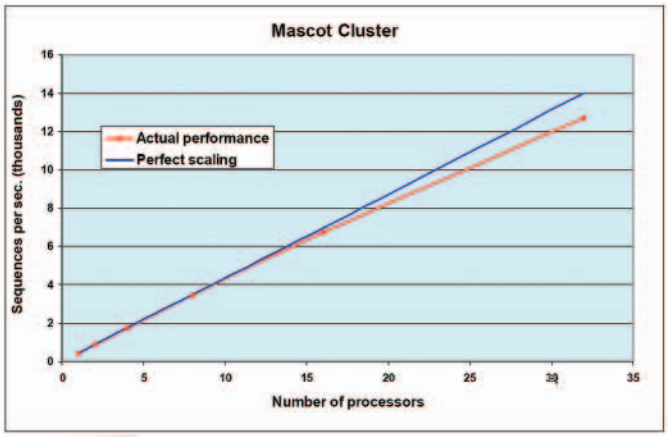
Mascot Daemon can take advantage of a variety of data import filters, including Mascot Distiller, to automate the processing of raw data files into peak lists. For example, Finnigan LCQ and LTQ raw files can be processed into peak lists using either the utility supplied with Xcalibur or Mascot Distiller.

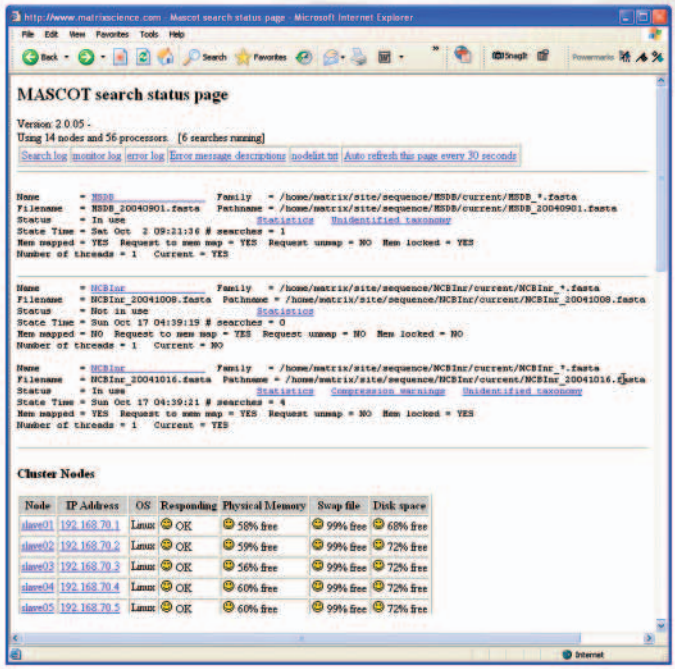
FAST, THREADED CODE, OPTIMIZED FOR A WIDE RANGE OF SINGLE AND MULTI-PROCESSOR PLATFORMS.

Mascot Server is available for all these popular computing platforms:

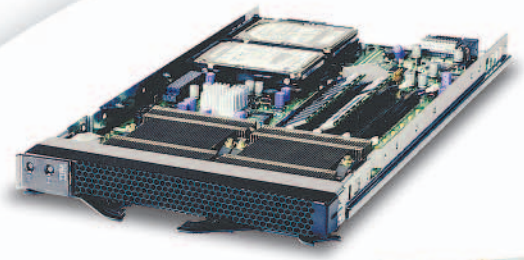
- Microsoft Windows NT / 2000 / XP / 2003
- Linux
- IBM AIX
- Solaris
- HP Tru64 Unix
- SGI IRIX

If required, Mascot can be executed on a multi-processor server or a cluster of servers connected by a standard LAN. Cluster-mode execution is a standard feature of the code, and is enabled whenever the license is for four or more processors. Throughput scales almost perfectly with the number of processors.





In addition to our software-only solutions, Mascot Cluster provides a complete turn-key system for high throughput protein identification. The IBM eServer™ BladeCenter™ hardware is easy to manage and cost effective. The operating system can be either Microsoft Windows or Linux. A single, fully populated BladeCenter contains 14 blades (28 processors). Up to 6 BladeCenters, (168 Mascot processors), can be mounted in a single industry-standard 19" rack. The BladeCenter chassis incorporates redundant power supplies and individual blades are hot-swappable.

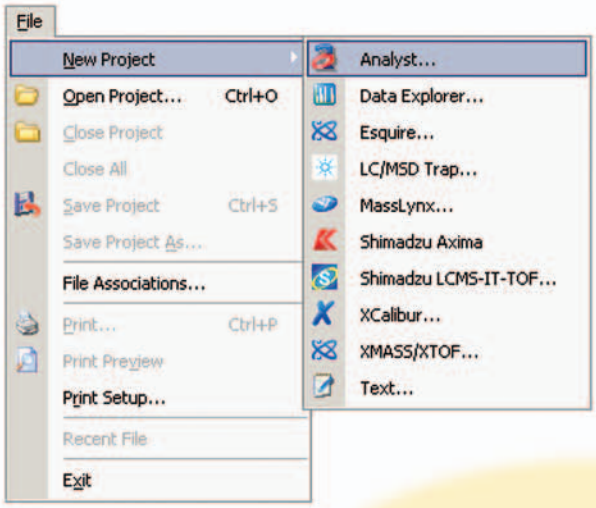


IBM Business Partner

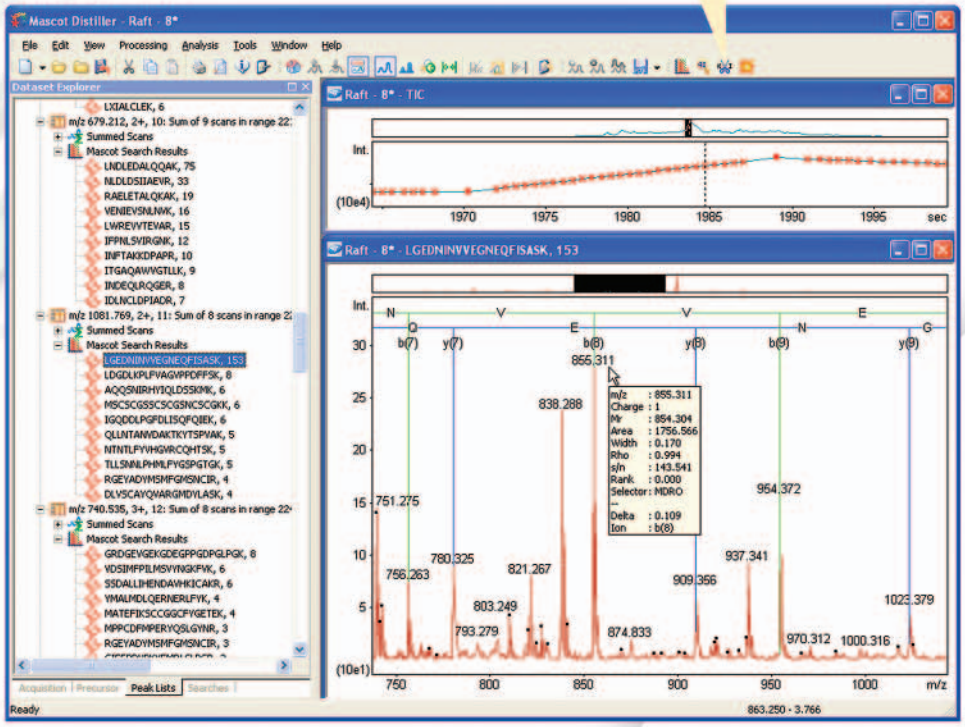
Mascot Distiller: Cross-platform browsing and processing of raw mass spectrometry data

Mascot Distiller processes raw mass spectrometry data into high quality peak lists for database searching. The graphical user interface provides a simple and intuitive data browser for viewing spectra and generating peak lists from a wide range of mass spectrometers. The processing code is in the form of a Windows COM library, that can be called by applications such as Mascot Daemon, or by programs you write yourself.

Mascot Distiller detects a peak by fitting an ideal isotopic distribution to the experimental data. The advantage of this approach is that the complete experimental distribution is fitted, not just the ^{12}C peak. The charge state is automatically determined and the peak list contains only monoisotopic peaks, even when the signal to noise is poor or the isotopic distribution not fully resolved. Smoothing is not necessary or desirable.



Mascot Distiller 2.0 includes a fast de novo sequencing algorithm, manual and automatic calling of sequence tags, *in silico* protein digestion and peptide fragmentation, and the ability to import and view Mascot search results.

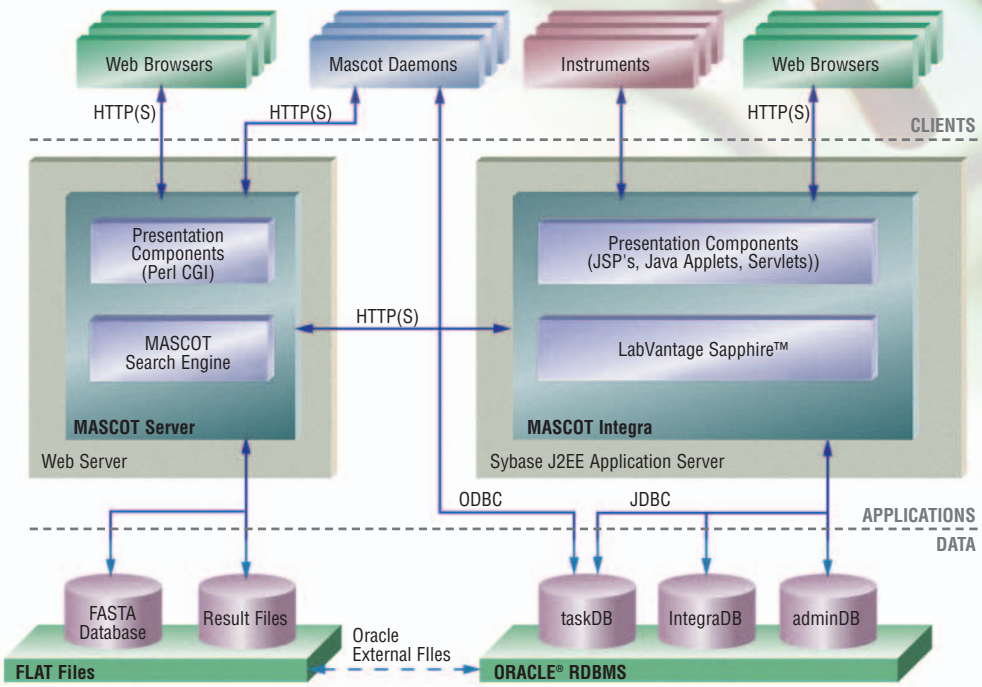


Mascot Integra: Data management for Proteomics

Mascot Integra is a complete solution for proteomics workflow automation and data mining.

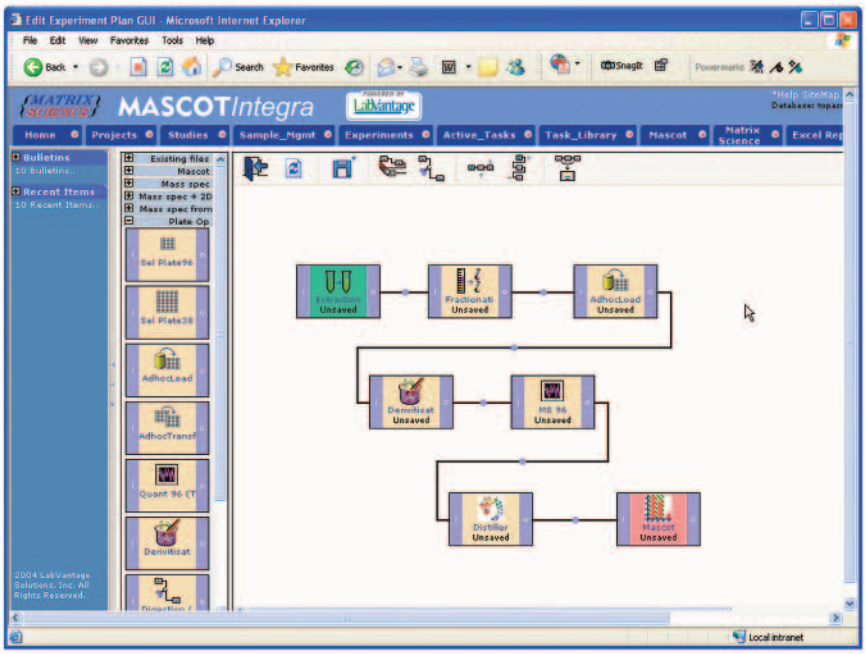
Unlike a traditional LIMS system, Mascot Integra is fully functional “out of the box’. Both gel-based and chromatography-based workflows can be modelled using an intuitive, graphical user interface. Sample and workflow management is powered by the technology of Sapphire™ LIMS from LabVantage Solutions Inc.

The database engine is Oracle® and the system is truly multi-tier, requiring only a web browser on each client. Mascot Integra automates Matrix Science's proven tools, including Mascot Distiller for browsing and processing mass spectrometry data, and the Mascot search engine for protein identification and characterization.



The database schema is structured to facilitate data mining, and Microsoft Excel provides a familiar interface for custom reports.

Mascot Integra will scale to the largest projects, yet has a very affordable entry level, making it a practical choice for smaller laboratories.

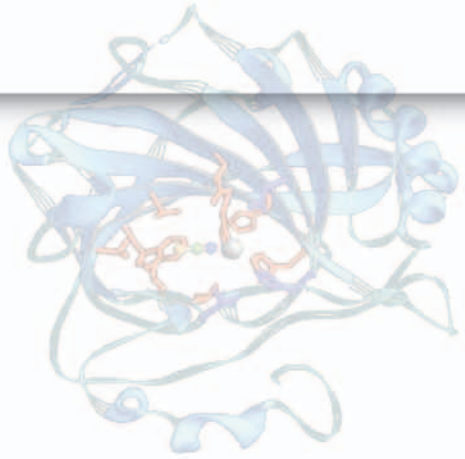


CFR21 part 11 ERES compliant

ORACLE PARTNER

POWERED BY LabVantage

All trade marks and service marks on this page are the properties of their respective owners and are hereby acknowledged



Contacts

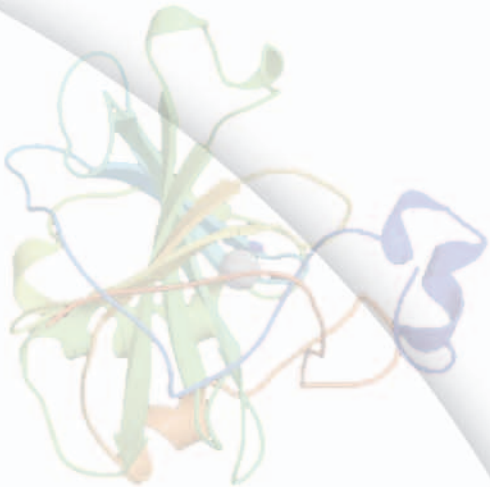
Headquarters:
Matrix Science Ltd.
8 Wyndham Place
London W1H 1PP
UK
Phone: +44 20 7723 2142
Fax: +44 20 7725 9360
Email: info@matrixscience.com

USA & Canada:
Matrix Science Inc.
225 Franklin Street, 26th Floor
Boston, MA 02110
USA
Phone: (800) 716 6702
Fax: (800) 716 6704
Email: info@matrixscience.com

Japan:
Matrix Science K.K.
KN-bldg 3F
6-10-12, Sotokanda
Chiyoda-ku, Tokyo 101-0021
Japan
Phone: 03 5807 7895
Fax: 03 5807 7896
Email: info-jp@matrixscience.com

About Matrix Science

Matrix Science is an independent bioinformatics company specialising in products and services for mass spectrometry and proteomics. For further information on all of our products and for free access to the Mascot search engine, visit the Matrix Science web site – http://www.matrixscience.com



MATRIX SCIENCE

www.matrixscience.com