Spectrum Mill MS Proteomics Workbench

Comprehensive tools for MS proteomics
Meeting the challenge of proteomics data analysis

Mass spectrometry is a core technology for proteomics research, but large-scale proteome research can generate enormous volumes of MS data. Turning this data into meaningful biological information can present a challenge every bit as great as acquiring the original data. Agilent’s Spectrum Mill MS proteomics workbench provides a comprehensive suite of software tools for processing MS and MS/MS spectra, determining protein identities and expression levels, and creating meaningful cross-sample and cross-experiment result summaries. Capabilities include:

- MS/MS spectral extraction and quality assessment tools to eliminate noise and poor-quality data from database searches
- Support for analyses from multiple instrument types in multiple data formats
- Database searching of either MS or MS/MS data with identification of mutations, post-translational modifications (PTMs), and chemical modifications
- Automatic and interactive results validation and re-searching using alternate parameters, modes (identity, homology, homology with PTMs), or databases
- Mass-based, semiquantitative analysis for first-order expression-level comparisons and also support for full ICAT quantitation
- Cross-experiment results summaries
- De novo sequencing of peptides from proteins not present in databases

These capabilities are carefully integrated into a powerful solution for MS proteomics data analysis.

Spectrum Mill workbench is web server based and accessed through a familiar web browser.

Agilent Technologies

Spectrum Mill MS Proteomics Workbench

Mass Spectral Interpretation Tools
- Data Extractor
- MS/MS Search
- Spectrum Matcher
- de novo Sequencing

Result Summary Tools
- ProteinPeptide Summary
- Spectrum Summary
- de novo Summary

Useful Tables:
- Mutation Mass Shifts
- Dipeptide Masses

Manuals:
- Quick Start Guide
- Application Guide
- Familiarization Guide

Utilities:
- Tool Belt
- Protein Databases
- Peptide Selector
- MS Digest
- MS Edman
- MS Product
- MS Comp
- MS Isotope

Help:
- Spectrum Mill Basics
- MS/MS Search
- PMF Search
- Protein Databases
- Server Administration

Collection of Spectrum Mill utility tools

Manipulate FASTA sequence databases for use with Spectrum Mill software
Select peptides from protein digest likely to produce high quality MS/MS spectra
Predict peptide masses for enzymatic digestion of protein
Search database with text or partial peptide sequence
Predict product ion masses from peptide sequence
Calculate AA compositions fitting parent mass and isotope patterns of peptides

Spectrum Mill workbench is web server based and accessed through a familiar web browser
Intelligent spectral extraction speeds protein identification

The speed and efficiency with which the Spectrum Mill workbench searches protein databases is due in large part to intelligent extraction and quality assessment of data before searching begins. The Spectrum Mill extraction software assesses MS/MS spectral quality based on sequence tag length and signal-to-noise criteria. Spectra that do not meet the criteria are excluded from the data set.

In order to further reduce data set size and search times, the extraction software:

- Eliminates duplicate spectra by merging scans from the same precursor within a specified time window
- De-isotopes and centroids the raw data
- Assigns charge states if possible

By excluding poor quality spectra and reducing unnecessary duplication, the speed of searching is greatly increased. This also reduces the number of false positives, which are common in some programs that do not use intelligent spectral extraction.

Data extractors for multiple instruments and data formats

Add-on data extractors are available for processing non-Agilent data formats:

- .RAW (Thermo Finnigan)
- .wiff (AB/MDS-SCIEX)
- .pkl (generic peak list - Waters/Micromass and others)

The extractor modules are optimized for the unique mass spectra characteristics of the type of mass spectrometer that generated the data.
Spectra from the remaining unvalidated matches can be re-searched using alternate parameters or databases. Iterative searches can be done only on spectra from unvalidated matches so that each subsequent search involves fewer spectra and is faster.

For spectra that do not produce a good database match but are of good quality, the reviewer has the option of applying a powerful de novo sequencing program (page 8).
Automatic match validation

The Spectrum Mill workbench features a unique autovalidation feature. Applying specified values for overall score and percent-scored peak intensity, the software evaluates database matches and marks those matches that are clearly correct. The spectra associated with these matches are isolated from spectra associated with unvalidated matches.

Interactive match validation

Data and spectra from unvalidated matches can then be reviewed and validated interactively by the reviewer. Matches the reviewer determines are good are added to the previously validated matches. Numerous features are included to make even manual spectral review fast and easy.

Autovalidation speeds data review by automatically validating the highest quality matches based on match score.
Complex data made accessible

One of the Spectrum Mill workbench’s most important benefits is its ability to summarize and correlate results — from single samples or complex studies — in ways that make the information accessible to biologists and biochemists as well as mass spectrometrists. The results of an MS/MS search, for example, can be displayed in multiple ways.

In peptide mode, a mass spectrometrist can review peptides matches in order of match quality, viewing peptide masses and mass spectra, protein identities and molecular weights, and even the relative abundances of the peptides.

Using protein mode, a biologist can review the same sample, viewing the protein names, match scores, percent coverages, database accession numbers, and relative abundances of the various proteins.

A user can also review multiple samples, comparing information such as expression levels. Large data sets can be compared across multiple experiments and the results can be readily summarized at the protein level.

Quantitative as well as qualitative information

Identities and mutations or modifications are only some of the information sought from proteomics studies. Equally critical are expression levels for the identified proteins. The Spectrum Mill workbench provides multiple tools for determining expression levels.

Peptide summary presents results, including annotated spectra, in an MS-centric format
**Fast, easy semiquantitative analysis**

The Spectrum Mill workbench includes a unique feature for comparing relative abundances of peptides and proteins. During spectral extraction, ion chromatograms are extracted for each peptide precursor ion and peak areas are determined. For each protein subsequently identified, the mean of the peak intensities of the component peptides provides a rough measure of the relative abundance of that protein in the sample. These intensities are displayed numerically and through color coding in the protein/peptide results summaries.

Although semiquantitative in nature, these comparisons of protein abundance based on extracted ion chromatograms (EICs) are sufficient to reveal two- to five-fold changes in relative concentrations.

**ICAT compatibility**

For applications where the semiquantitative approach is not sufficiently accurate, isotope-coded affinity tags (ICAT) can be used to achieve more accurate quantitation. The Spectrum Mill workbench fully supports ICAT analyses. Cys modifications include D0/D8, 12C/13C, or mixed labels.

---

<table>
<thead>
<tr>
<th>Protein Name</th>
<th>Database Accession</th>
<th>%AA Coverage</th>
<th>Distinct Peptides</th>
<th>Distinct Summed M</th>
<th>Group #</th>
<th>Protein MW (Da)</th>
<th>[M+H]$^+$ mass intensity</th>
<th>[M+H]+ mass intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>[A]</td>
<td>[A]</td>
<td>[A]</td>
<td>[A]</td>
<td>[A]</td>
<td>[A]</td>
<td>[A]</td>
<td>[A]</td>
<td>[A]</td>
</tr>
<tr>
<td>[B]</td>
<td>[B]</td>
<td>[B]</td>
<td>[B]</td>
<td>[B]</td>
<td>[B]</td>
<td>[B]</td>
<td>[B]</td>
<td>[B]</td>
</tr>
<tr>
<td>[C]</td>
<td>[C]</td>
<td>[C]</td>
<td>[C]</td>
<td>[C]</td>
<td>[C]</td>
<td>[C]</td>
<td>[C]</td>
<td>[C]</td>
</tr>
<tr>
<td>[D]</td>
<td>[D]</td>
<td>[D]</td>
<td>[D]</td>
<td>[D]</td>
<td>[D]</td>
<td>[D]</td>
<td>[D]</td>
<td>[D]</td>
</tr>
<tr>
<td>[E]</td>
<td>[E]</td>
<td>[E]</td>
<td>[E]</td>
<td>[E]</td>
<td>[E]</td>
<td>[E]</td>
<td>[E]</td>
<td>[E]</td>
</tr>
<tr>
<td>[F]</td>
<td>[F]</td>
<td>[F]</td>
<td>[F]</td>
<td>[F]</td>
<td>[F]</td>
<td>[F]</td>
<td>[F]</td>
<td>[F]</td>
</tr>
<tr>
<td>[G]</td>
<td>[G]</td>
<td>[G]</td>
<td>[G]</td>
<td>[G]</td>
<td>[G]</td>
<td>[G]</td>
<td>[G]</td>
<td>[G]</td>
</tr>
<tr>
<td>[H]</td>
<td>[H]</td>
<td>[H]</td>
<td>[H]</td>
<td>[H]</td>
<td>[H]</td>
<td>[H]</td>
<td>[H]</td>
<td>[H]</td>
</tr>
<tr>
<td>[I]</td>
<td>[I]</td>
<td>[I]</td>
<td>[I]</td>
<td>[I]</td>
<td>[I]</td>
<td>[I]</td>
<td>[I]</td>
<td>[I]</td>
</tr>
<tr>
<td>[J]</td>
<td>[J]</td>
<td>[J]</td>
<td>[J]</td>
<td>[J]</td>
<td>[J]</td>
<td>[J]</td>
<td>[J]</td>
<td>[J]</td>
</tr>
</tbody>
</table>

Protein-protein summary allows comparison of data, including color-coded relative abundance information, from multiple samples.
De novo spectral interpretation for peptides not matched in any database

Proteins not found in any database can be a significant barrier to deeper understanding of proteomes. Manual spectral interpretation requires extensive knowledge and practice; it can be very challenging and time consuming. De novo sequencing using various mathematical algorithms has been applied with varying degrees of success. The major challenge of de novo MS/MS spectral interpretation arises from the fact that peptides do not evenly fragment between each amino acid, resulting in incomplete spectral information.

The Spectrum Mill workbench employs the Sherenga de novo sequencing algorithm which uses advanced graph theory to discard unrealistic solutions and generate a ranked list of potential peptide sequences. The Spectrum Mill de novo sequencing program compensates for common spectral difficulties such as noise and incomplete fragmentation. It also employs rigorous scoring routines with both positive scoring for anticipated ions and negative scoring for uncharacteristic spectral patterns.

The Spectrum Mill de novo sequencing result review enables simultaneous viewing of both the database search results and the de novo results.

De novo spectral interpretation results with results table and MS/MS spectra annotated with fragments corresponding to proposed interpretation

For more information

For more information about the Agilent Spectrum Mill MS proteomics workbench, call toll free:

1-800-227-9770 (in the U.S. and Canada)

In other countries, please call your local Agilent Technologies life sciences and chemical analysis sales office or authorized Agilent Technologies distributor.

You can also visit our site on the World Wide Web at:

www.agilent.com/chem