Importing Data from Different Results Formats into MS-Viewer

Peter R. Baker[1], Alma L. Burlingame[1] and Robert J. Chalkley[1]

[1]Mass Spectrometry Facility, Dept. of Pharmaceutical Chemistry, University of California, San Francisco, USA

Introduction

MS-Viewer is a spectral viewer that allows the annotated spectra from database search results to be viewed interactively. It requires one or more peak list files in one of a range of common formats and results in a spreadsheet file. Hence, results from almost any search engine can be displayed with little work. Once uploaded other internet users can access the data via a keyword. Thus the web site can act as a repository for published results or as a format for sharing results with colleagues or collaborators. The software can handle peptides with ambiguous PTM site assignments and cross-linked peptides. It also supports visual comparison of two different assignments to the same spectrum.

Recently the program has been extended to allow it to read results in Pride XML, pepXML and Thermo msf formats. This means that over half of the data sets in ProteomeXchange can potentially be displayed in MS-Viewer. As a demonstration, data from five of these data sets can now be accessed directly from the Protein Prospector public web site.

How to Use

The program input is either a single peak list file or an archive of peak list files and a database search results file. The results file is expected to be in tabular form (tab-delimited text or comma separated value file) with columns including peptide sequences, spectrum identifiers and precursor charges. A fraction column is also required if multiple peak lists are uploaded. An arbitrary number of other columns containing other information may also be present. Alternatively a number of XML and database formats are supported. Individual spectra in the report can be viewed by clicking on entries in the peptide column. There is also a facility for re-searching individual spectra using the MS-Tag program. Data from up to four columns can be used for sorting the table. Reports can be saved and then viewed in the future by other web users who enter a 10 digit randomly generated code.

Users can write their own Perl scripts for automatic results file conversion.

Another feature of the spectral viewer is the ability to deal with ambiguous site assignments for modified amino acids and cross-linked peptides.

Peak list formats supported: mgf, mzData, pkl, dta, mzML, mzXML. Also ms2 coming soon.

Results file formats supported: Protein Prospector/X!Tandem Tab Delimited Text, Mascot CSV, tabular file with arbitrary number of columns. Also pepXML, Thermo MSF, PRIDEXML coming soon.

Archive and compression formats supported: zip, rar, 7z, gz, z, bz2, cmn, tar, tgz, tar.gz, taz, tar.z.

Conclusion

MS-Viewer is available at http://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msviewer
or for local installation by emailing ppadmin@cgl.ucsf.edu.

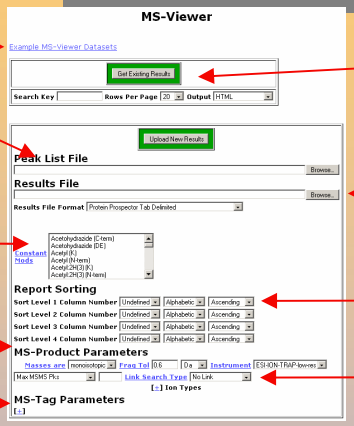
The example MS-Viewer datasets are at
http://prospector.ucsf.edu/prospector/html/misc/viewereg.htm

A video explaining MS-Viewer is available at http://vimeo.com/30462677.

Acknowledgements

This work was supported by NIH NIGMS 8P41GM103481, 1S10RR019934 and the Biotechnology and Biological Sciences Research Council of the UK.

MS-Viewer Form



Link to example datasets.

The peak list file can either be a single file or a file archive (zip, rar, 7z, etc) containing multiple peak list files. Most of the common peak list formats (mgf, mzXML, mzML, dta, pkl, ms2, etc) are supported.

Constant amino acid modifications need to be specified if they are not present in the relevant report column.

Parameters for the spectral viewer.

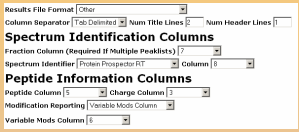
Default parameters for MS-Tag search.

Once a report has been formatted it can be saved and subsequently recalled by another web user entering a 10 digit code.

The results file can either be a Protein Prospector or X!Tandem Tab Delimited Text file, a Mascot CSV file, a PRIDE XML file, a pepXML file, a Thermo msf file or a tabular file with an arbitrary number of columns (see right).

Reports can be sorted based on the contents of up to 4 columns.

For displaying cross-linked results.



To deal with other results file formats the user needs to specify the columns containing the peptide hits, the spectral identifiers and the precursor charges.

Example MS-Viewer Datasets