Improvements to Protein Prospector's MS-Viewer

Peter R. Baker[1] and Robert J. Chalkley[1]

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

Introduction

The ability to share results and make annotated spectra available is important to many researchers and is a prerequisite to publish in some proteomic journals. A spectral viewer called MS-Viewer[1] is available as part of Protein Prospector that allows the annotated spectra from database search results produced by most search engines to be viewed interactively. Users can upload the results and associated peak list files in a wide variety of formats. Once uploaded other internet users can access the data via a keyword. The web site has been used as a repository for results from over a dozen different formats published in seven different journals. The poster will outline recent improvements to MS-Viewer.

Supported Formats

Protein Prospector
Protein Prospector Cross-Linked Peptides
Mascot CSV
Protein Pilot
Spectrum Mill
X!Tandem Tab-delimited
PRIDE XML
pepXML (TPP)
MaxQuant
Thermo MSF (Proteome Discoverer)
MS-Amanda
Morpheus
Blib (Bibliospec/Skyline)
SpectraST sptxt
NIST MSP

In addition, any results format that contains columns including peptide sequence with modifications; charge state; spectrum identifier such as scan number; and corresponding peak list file name, can be imported with minimal editing.

Results Filtering

Very large datasets, with a lot of redundant identifications, are often produced using the latest instruments. These results can be automatically filtered during upload to MS-Viewer to only retain the most confident identification of a given peptide species. An example (from a MaxQuant data set) can viewed by entering the key kjfmg24iwf into MS-Viewer[2].

In this example 1.3 million spectral identifications were consolidated to 156,000 unique peptide identifications, reducing starting peak list and results files from 14 GB to 900 MB. This data set was processed through the web interface in just under 90 minutes: 21 minutes for upload of zipped peak list files and the results file, 59 minutes for unpacking and conversion of .apl and msms.txt results files into a standard format, then 4 minutes to filter results.

Summary

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

The MS-Viewer published dataset brower is available at http://msviewer.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msviewrep.

Videos explaining MS-Viewer are available at http://vimeo.com/30462677 and http://vimeo.com/ 112225814.

References

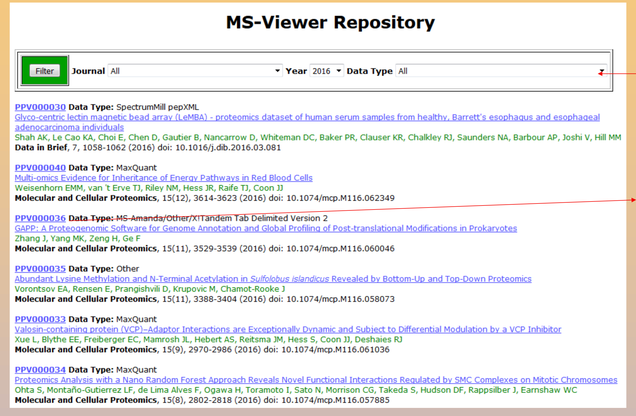
1. Baker PR and Chalkley RJ, MS-Viewer: A Web Based Spectral Viewer for Proteomics Results. Molecular and Cellular Proteomics, doi:10.1074/mcp.O113.037200 (2014)
2. Tatham MH, Cole C, Scullion P, Wilkie R, Westwood NJ, Stark LA, Hay RT, A proteomic approach to analyse the aspirin-mediated lysine acetylome. Molecular and Cellular Proteomics, 16(2), 310-326, doi: 10.1074/mcp.O116.065219 (2017)
3. Picotti P et al, A complete mass-spectrometric map of the yeast proteome applied to quantitative trait analysis. Nature, 494, 266-270, doi: 10.1038/nature11835 (2013)

Acknowledgements

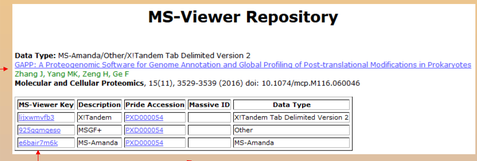
This work was supported by NIH grant 8P41GM103481 and the Dr Miriam and Sheldon G. Adelson Medical Research Foundation.

MS-Viewer Published Dataset Browser

The repository currently houses public data from about 40 publications. These datasets can be viewed and downloaded using a dataset browser.



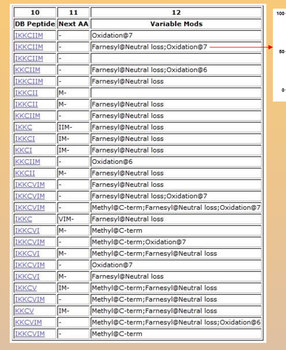
Published datasets can be filtered by journal, year or data format



Links to MS-Viewer datasets associated with the published paper.

Links to PRIDE and/or Massive submissions

Improved Filtering Options

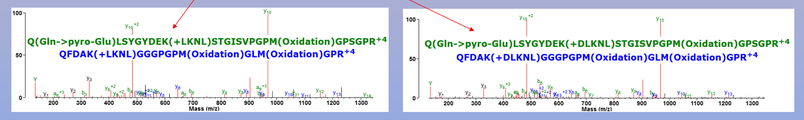


Column Filtering Options

Results can be sorted or filtered by column content, including setting minimum scores or maximum expectation value thresholds.

Support for Multiple Crosslinker Data

3	5	8	9	11	12	14	15	16	17	18	19	33	36
ppm	Crosslinked Peptide	Fraction	RT	Score	Score Diff	Sc 1	Exp 1	Rk 1	Sc 2	Exp 2	Rk 2	Protein Name	Protein Name
-9.5	Q(Gln->pyro-Glu)LSYGYDEK(+DeHLNL)STGISVPGPM(Oxidation)GPSGPR QFDAK(+DeHLNL)GGGPGPM(Oxidation)GLM(Oxidation)GPR	P20161009-07	90.567	83.6	33.8	49.8	4.0e-10	1	36.6	2.5e-7	13	CO1A1	CO1A2
-3.6	Q(Gln->pyro-Glu)LSYGYDEK(+DLKNL)STGISVPGPM(Oxidation)GPSGPR QFDAK(+DLKNL)GGGPGPM(Oxidation)GLM(Oxidation)GPR	P20161009-05	83.733	77.8	26.7	51.1	3.3e-11	4	29.2	3.3e-6	16	CO1A1	CO1A2
-8.0	Q(Gln->pyro-Glu)LSYGYDEK(+LKNL)STGISVPGPM(Oxidation)GPSGPR QFDAK(+LKNL)GGGPGPM(Oxidation)GLM(Oxidation)GPR	P20161009-07	86.650	72.7	26.1	46.6	3.6e-11	2	25.9	2.0e-6	14	CO1A1	CO1A2
-11	Q(Gln->pyro-Glu)LSYGYDEK(+HLKNL)STGISVPGPM(Oxidation)GPSGPR QFDAK(+HLKNL)GGGPGPM(Oxidation)GLM(Oxidation)GPR	P20161009-07	83.850	79.7	25.0	54.7	1.1e-11	3	27.5	9.7e-6	15	CO1A1	CO1A2



MS-Viewer can now display results from a search considering 4 different cross-linkers.

Example of SpectraST sptxt File

1	2	3	4	5	6	7	8	9	10	11
Prev AA	DB Peptide	Next AA	Modifications	Precursor MZ	Precursor Charge	Spectrum	Num Peaks	Peptide Type	Inst	Protein
R	AAAAODEITGDGTTTVVCLVGELLR	Q	Carbamidomethyl@18	1280.6498	2	27	47	Tryptic	3/FTMS,3,0/ITMS,1,0/it,6,0	1/YDR188W
R	AAAAODEITGDGTTTVVCLVGELLR	Q	Carbamidomethyl@18	854.1023	3	28	48	Tryptic	3/FTMS,8,0/ITMS,1,0/it,4,0	1/YDR188W
K	AAAEGPMK	G	Oxidation@7	395.6918	2	39	77	Tryptic	1/it,12,0	1/YJL052W
K	AAAEGVANLHLDEATGEMVSK	S	Oxidation@18	1065.0126	2	43	85	Tryptic	2/FTMS,4,0/it,21,0	1/YDR037W
K	AAAEGVANLHLDEATGEMVSK	S	Oxidation@18	710.3441	3	45	106	Tryptic	3/FTMS,16,0/ITMS,1,0/it,28,0	1/YDR037W
R	AAAEVLYQYILMQADLDVIR	K	Oxidation@12	771.0716	3	52	81	Tryptic	1/FTMS,8,0	1/YKL057C
T	AAAFGNCHGLYAQDIALRPEILAEHQK	Y	Carbamidomethyl@7	974.8275	3	58	80	Semi- tryp	2/ITMS,1,0/it,5,0	1/YKL060C
R	AAALSIIGTAVQNNLDSQNNFMK	Y	Oxidation@22	812.7444	3	94	150	Tryptic	1/FTMS,1	1/YBR101C

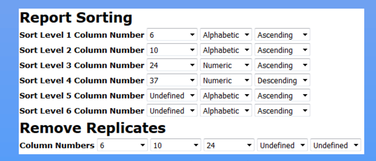
MS-Viewer is able to display and annotate the contents of the ISB Yeast Orbitrap SRM atlas[3]

Display of Very Large MaxQuant Datasets

1	3	6	10	22	24	27	30	32	34	37	62
Raw file	Scan number	Sequence	Variable mods	Gene Names	Charge	Type	m/z	Mass Error [ppm]	Retention time	Score	Protein group IDs
20150225_MT_Crude_MixE	63009	AAAAAAAAAAGAAGGR	Acetyl@N-term	PABPN1	2	MULTI-MSMS	620.82327	-0.46894	134.27	159.08	3615
20150225_MT_Crude_MixG	63041	AAAAAAAAAAGAAGGR	Acetyl@N-term;Label:13C(6)15N(4)@16	PABPN1	2	ISO-MSMS	625.82741	-0.98176	134.3	178.15	3615
20150225_MT_Crude_MixF	62942	AAAAAAAAAAGAAGGR	Acetyl@N-term;Label:13C(6)@16	PABPN1	2	MULTI-MSMS	623.83334	-0.65586	134.36	147.18	3615
20150225_MT_Crude_MixB	58550	AAAAAAAAAPAAAATAPTTAATTAATAAQ		SRP14	2	ISO-MSMS	1184.6088	0.45686	126.5	112.88	2397
20150225_MT_Crude_MixF	58444	AAAAAAAAAPAAAATAPTTAATTAATAAQ		SRP14	3	ISO-MSMS	790.07495	0.6601	126.08	66.853	2397
20150424_MT_Crude_MixB_240_g2	95603	AAAAAAAGDSDSWDADAFSVEDPVRK	Acetyl@N-term	EIF3J	3	MULTI-MSMS	879.06841	-1.2002	191.75	99.366	1770
20150424_MT_Crude_MixA_240_g3	25282	AAAAAAAGDSDSWDADAFSVEDPVRK	Acetyl@N-term;Label:13C(6)15N(4)@25;Label:13C(6)15N(2)@26	EIF3J	3	ISO-MSMS	885.0759	-0.64648	54.865	105.79	1770
20150424_MT_Crude_MixA_240_g3	25295	AAAAAAAGDSDSWDADAFSVEDPVRK	Acetyl@N-term;Label:13C(6)@25;Label:2H(4)@26	EIF3J	3	ISO-MSMS	882.41682	3.5217	54.891	88.79	1770
20150225_MT_Crude_MixF	14377	AAAAAAALQAK		RPL4	2	MULTI-MSMS	478.77982	0.067774	44.352	122.26	837
20150225_MT_Crude_MixE	14601	AAAAAAALQAK	Label:13C(6)15N(2)@11	RPL4	2	MULTI-MSMS	482.78692	0.19452	44.26	122.26	837
20150225_MT_Crude_MixF	14299	AAAAAAALQAK	Label:2H(4)@11	RPL4	2	MULTI-MSMS	480.79237	0.067774	44.195	110.14	837
20150225_MT_Crude_MixD	79582	AAAAAAGAASGLPGPVAQGLK	Acetyl@N-term	IPO9	2	ISO-MSMS	895.9916	-0.11152	165.31	135.28	4053
20150225_MT_Crude_MixF	79682	AAAAAAGAASGLPGPVAQGLK	Acetyl@N-term	IPO9	3	ISO-MSMS	597.66349	0.18509	165.17	84.829	4053
20150225_MT_Crude_MixC	79257	AAAAAAGAASGLPGPVAQGLK	Acetyl@N-term;Label:13C(6)15N(2)@21	IPO9	2	MULTI-MSMS	899.9987	0.65778	164.94	180.92	4053
20150225_MT_Crude_MixC	79245	AAAAAAGAASGLPGPVAQGLK	Acetyl@N-term;Label:13C(6)15N(2)@21	IPO9	3	MSMS	600.33489	NaN	164.92	80.638	4053
20150225_MT_Crude_MixB	79635	AAAAAAGAASGLPGPVAQGLK	Acetyl@N-term;Label:2H(4)@21	IPO9	2	MULTI-MSMS	898.00415	-0.1807	165.33	157.12	4053
20150225_MT_Crude_MixB	79662	AAAAAAGAASGLPGPVAQGLK	Acetyl@N-term;Label:2H(4)@21	IPO9	3	MSMS	599.00519	NaN	165.37	101.15	4053

Most of the 68 columns have been hidden.

SILAC labels are inserted by the conversion script.



Report Sorting and Remove Replicate options facilitate removal of redundant identifications

MS-Viewer can display very large MaxQuant data sets that the MaxQuant Viewer often fails to open. It is able to filter results such that only the best scoring hit for a given combination of peptide sequence, variable mods and precursor charge is retained.