

Sharing Annotated Spectra when Standard Formats are not an Option

Robert J. Chalkley and Peter R. Baker

Mass Spectrometry and Proteomics Resource, University of California San Francisco, USA



Introduction

The requirement by journals for authors to make available annotated proteomic results has been driven by the recognition that authors cannot look at all of their reported results, so by sharing data associated with a manuscript readers can make an independent assessment and reanalysis of results of particular interest. This process has been heavily facilitated by the development of standard file formats such as mzML, mzIdentML and mzTab. However, for some search engines or results types, outputs in these formats are not an option, presenting a challenge for authors to comply with journal guidelines.

MS-Viewer, part of the free Protein Prospector package, can produce annotated spectra from many search engine outputs, including simple tab-delimited text files. This allows it to annotate data from search engines that do not support standard formats. It can also annotate cross-linked peptide identifications. A website allows submission of results to a public repository, and there is a searchable browser that allows one to see all submissions associated with publications. There are links to associated proteomeXchange submissions, and PRIDE provides links from datasets in their repository to MS-Viewer.

MS-Viewer has been used to publish results associated with about fifty publications in eleven different journals, and is also used for sharing results between laboratories. It is particularly popular with MaxQuant users, as it allows production of annotated spectra from much smaller file uploads than through using the MaxQuant Viewer and can handle large datasets that the MaxQuant Viewer struggles to display. Working with authors, we have successfully produced annotated spectra from all submissions.

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 be viewed by entering the key kjfmg24iwf into MS-Viewer¹.

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.

MS-Viewer Published Dataset Browser

The repository currently houses public data from about forty publications. These datasets can be searched, viewed and downloaded using the dataset browser.

MS-Viewer Repository

Filter Journal All Year 2016 Data Type All

PPV000030 Data Type: SpectrumMill pepXML
Glyco-centric lectin magnetic bead array (L-498A) - proteomics dataset of human serum samples from healthy, Barrett's esophagus and esophageal adenocarcinoma individuals
Shah AK, Le Cao KA, Choi E, Chen D, Gaudier B, Nancarrow D, Whiteman DC, Baker PR, Clauser KR, Chalkley RJ, Saunders NA, Barbour AP, Joshi V, Hill MM
Data in Brief, 7, 1059-1062 (2016) doi: 10.1016/j.dib.2016.03.081

PPV000040 Data Type: MaxQuant
Multi-omics Evidence for Involvement of Energy Pathways in Red Blood Cells
Weisenborn SM, van T Erve TJ, Riley NM, Hess JR, Raife T, Coon JJ
Molecular and Cellular Proteomics, 15(12), 3614-3623 (2016) doi: 10.1074/mcp.M116.062349

PPV000036 Data Type: MS-Amanda/Other/X!Tandem Tab Delimited Version 2
GAPP: A Proteogenomic Software for Genome Annotation and Global Profiling of Post-translational Modifications in Prokaryotes
Zhang J, Yang M, Zeng H, Ge F
Molecular and Cellular Proteomics, 15(11), 3529-3539 (2016) doi: 10.1074/mcp.M116.060046

PPV000035 Data Type: Other
Abundant lysine modifications and N-terminal acetylation in *Sulfolobus islandicus* revealed by bottom-up and top-down proteomics
Vorontsov EA, Rensen E, Prangishvili D, Krupovic M, Chamot-Rooke J
Molecular and Cellular Proteomics, 15(11), 3388-3404 (2016) doi: 10.1074/mcp.M116.058073

PPV000033 Data Type: MaxQuant
Valosin-containing protein (VCP) adaptor interactions are exceptionally dynamic and subject to differential modulation by a VCP inhibitor
Xue L, Blythe EC, Freiburger EC, Mamrosh JL, Hebert AS, Raitama JM, Hess S, Coon JJ, Deshaies RJ
Molecular and Cellular Proteomics, 15(9), 2970-2986 (2016) doi: 10.1074/mcp.M116.061036

PPV000034 Data Type: MaxQuant
Proteomics Analysis with a Nano Random Forest Approach Reveals Novel Functional Interactions Regulated by SMC Complexes on Mitotic Chromosomes
Ohta S, Montaño-Guerrero LF, de Lima Alves F, Ogawa H, Teramoto J, Sato N, Morrison CG, Takeda S, Hudson DF, Rappalber J, Earnshaw WC
Molecular and Cellular Proteomics, 15(8), 2802-2818 (2016) doi: 10.1074/mcp.M116.057885

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

MS-Viewer Repository

MS-Viewer Key	Description	PRIDE Accession	Massive ID	Data Type
isocm0b3	X!Tandem	EX0200054		X!Tandem Tab Delimited Version 2
isocm0e0	MSGF+	EX0200054		Other
isobar7ind6	MS-Amanda	EX0200054		MS-Amanda

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

Links to PRIDE and/or Massive submissions

Displaying Very Large MaxQuant Datasets

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. It is also able to translate MaxQuant PTM site localization scoring such that MS-Viewer can display localization ambiguity on annotated spectra.

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	AAAAAAAAAAGAGG	Acetyl@N-term	PABPN1	2	MULTI-MSMS	620.82227	-0.46894	134.27	159.08	3615
20150225_MT_Crude_MixG	63041	AAAAAAAAAAGAGG	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	AAAAAAAAAAGAGG	Acetyl@N-term;Label:13C(6)15N(4)@16	PABPN1	2	MULTI-MSMS	623.83334	-0.65586	134.36	147.18	3615
20150225_MT_Crude_MixB	58550	AAAAAAAAAAGAGG		SRP14	2	ISO-MSMS	1184.6088	0.45686	126.5	112.88	2397
20150225_MT_Crude_MixF	58444	AAAAAAAAAAGAGG		SRP14	3	ISO-MSMS	790.07495	0.6601	126.08	66.853	2397
20150424_MT_Crude_MixB_240_g2	95603	AAAAAAAAAGSDSDWADAFSDVSRVYK	Acetyl@N-term	E1F31	3	MULTI-MSMS	879.06841	-1.2002	191.75	99.366	1770
20150424_MT_Crude_MixA_240_g3	25282	AAAAAAAAAGSDSDWADAFSDVSRVYK	Acetyl@N-term;Label:13C(6)15N(4)@25;Label:13C(6)15N(2)@26	E1F31	3	ISO-MSMS	885.0759	-0.64648	54.865	105.79	1770
20150424_MT_Crude_MixA_240_g3	25295	AAAAAAAAAGSDSDWADAFSDVSRVYK	Acetyl@N-term;Label:13C(6)15N(4)@25;Label:2H(4)@26	E1F31	3	ISO-MSMS	882.41682	3.5217	54.891	88.79	1770
20150225_MT_Crude_MixF	14377	AAAAAAAAALQK		RPL4	2	MULTI-MSMS	1478.77982	0.067774	44.352	122.26	637
20150225_MT_Crude_MixE	14601	AAAAAAAAALQK	Label:13C(6)15N(2)@11	RPL4	2	MULTI-MSMS	1482.78692	0.19452	44.26	122.26	637
20150225_MT_Crude_MixF	14399	AAAAAAAAALQK	Label:2H(4)@11	RPL4	2	MULTI-MSMS	1480.79237	0.067774	44.195	110.14	637
20150225_MT_Crude_MixD	79582	AAAAAAAAAGASGLPQVAGQK	Acetyl@N-term	IP09	2	ISO-MSMS	895.9916	-0.11152	165.31	135.28	4053
20150225_MT_Crude_MixF	79582	AAAAAAAAAGASGLPQVAGQK	Acetyl@N-term	IP09	3	ISO-MSMS	597.66349	0.18509	165.17	84.829	4053
20150225_MT_Crude_MixC	79237	AAAAAAAAAGASGLPQVAGQK	Acetyl@N-term;Label:13C(6)15N(2)@21	IP09	2	MULTI-MSMS	899.9987	0.65778	164.94	180.92	4053
20150225_MT_Crude_MixC	79235	AAAAAAAAAGASGLPQVAGQK	Acetyl@N-term;Label:13C(6)15N(2)@21	IP09	3	MSMS	600.33489	NaN	164.92	80.638	4053
20150225_MT_Crude_MixB	79635	AAAAAAAAAGASGLPQVAGQK	Acetyl@N-term;Label:2H(4)@21	IP09	2	MULTI-MSMS	898.00415	-0.1807	165.33	157.12	4053
20150225_MT_Crude_MixB	79662	AAAAAAAAAGASGLPQVAGQK	Acetyl@N-term;Label:2H(4)@21	IP09	3	MSMS	599.00519	NaN	165.37	101.15	4053

SILAC labels are inserted by the conversion script.

MS-Viewer allows columns from the uploaded results to be hidden. In the above example, most of the 68 columns from the MaxQuant results file have been hidden (note the column numbering at the top of each column).

Uploaded results can be sorted or filtered by column content; e.g.

- Filter results to require a minimum score or maximum expectation value threshold
- Filter results to only phosphopeptides
- Filter results only to protein of interest

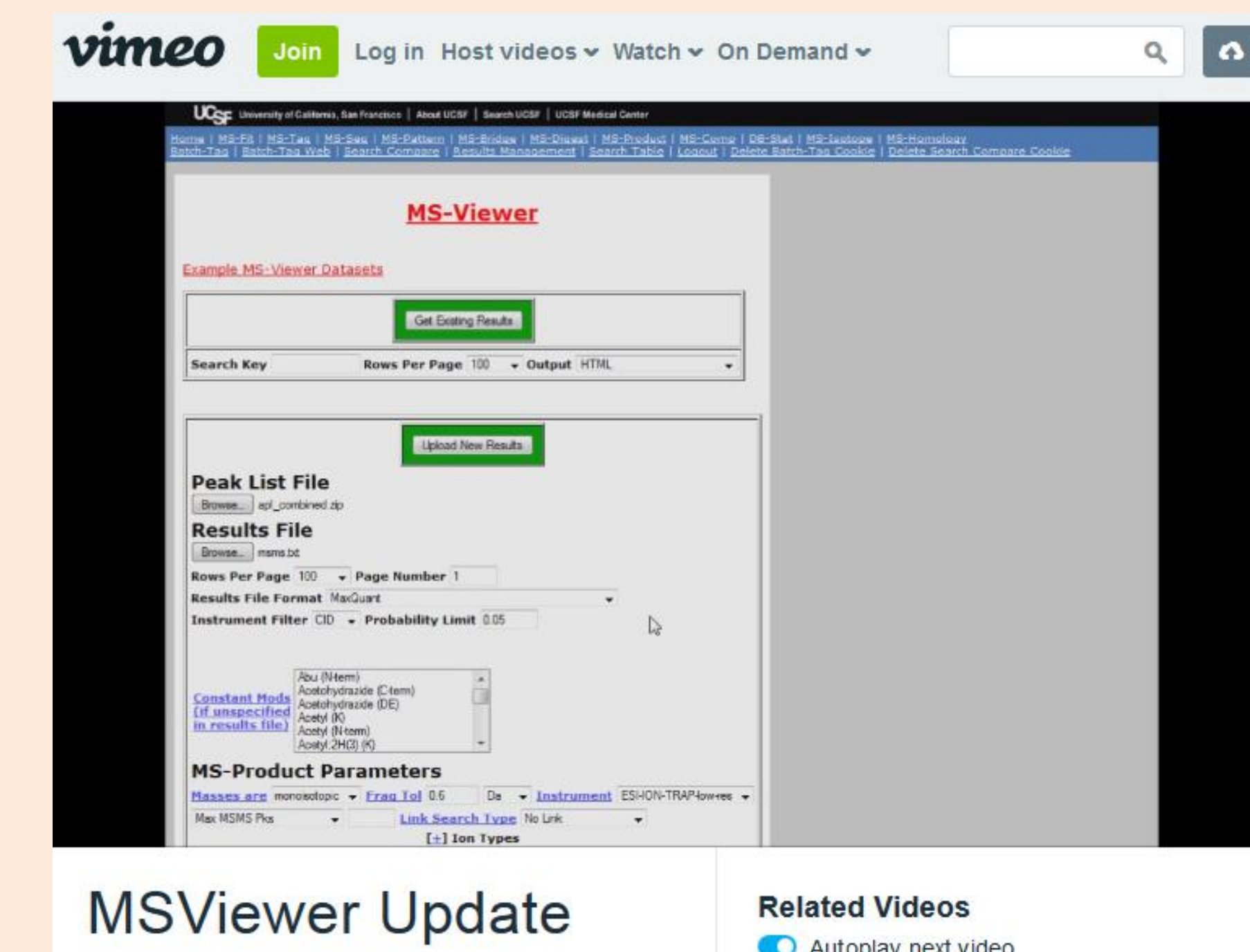
Report Filtering
Filter 1 Column Number Undefined Equals
Filter 2 Column Number Undefined Equals

Report Sorting
Sort Level 1 Column Number 6 Alphabetical Ascending
Sort Level 2 Column Number 10 Alphabetical Ascending
Sort Level 3 Column Number 24 Numeric Ascending
Sort Level 4 Column Number Undefined Numeric Descending
Sort Level 5 Column Number Undefined Alphabetical Ascending
Sort Level 6 Column Number Undefined Alphabetical Ascending

Remove Replicates
Column Numbers 6 10 24 Undefined Undefined

A report can be sorted by several columns then replicate identifications can be removed.

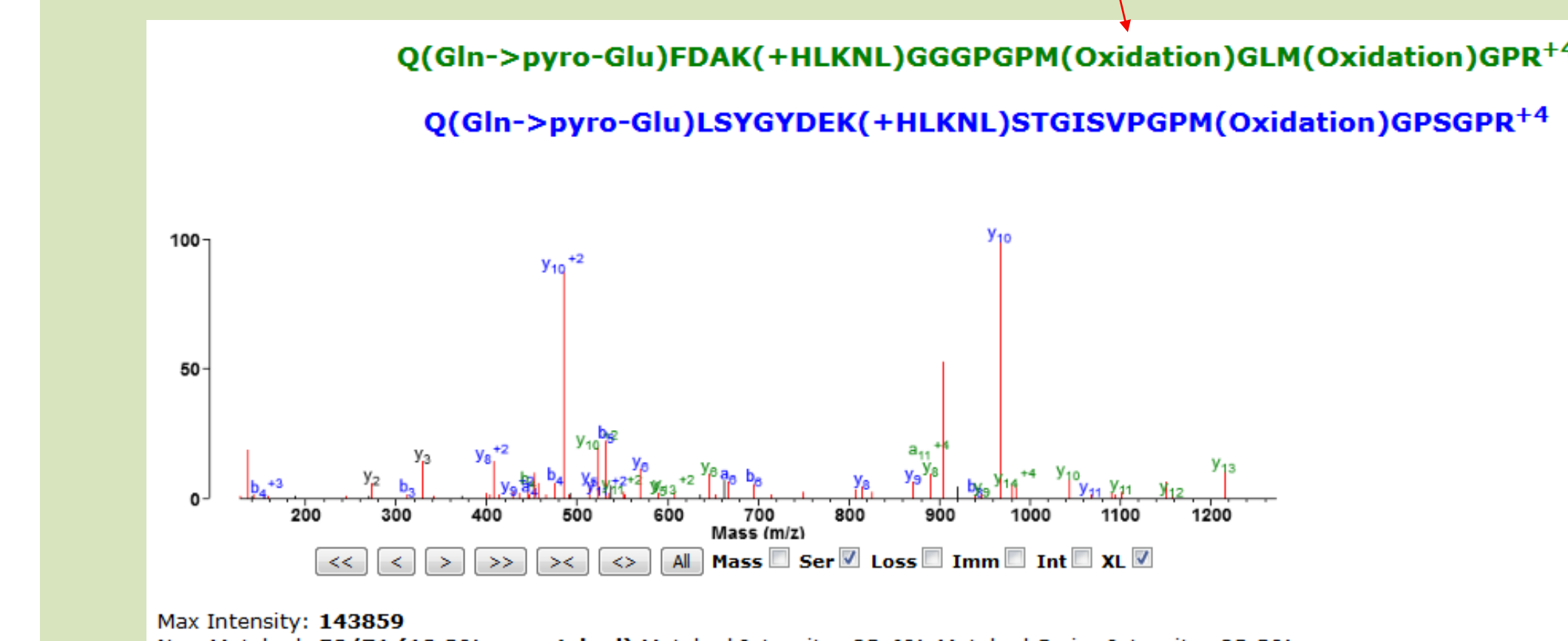
Video Tutorials



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

Displaying Cross-linked Peptide Results

1	2	3	5	9	11	12	13
m/z	z	ppm	Crosslinked Peptide	RT	Score	Score Diff	Expect
1095.4992	4	1.3	Q(Gln->pyro-Glu)FDAK(+HLKNL)GGGPGPM(Oxidation)GLM(Oxidation)GPR	50.183	85.6	40.4	2.1e-20
1095.4992	4	1.3	Q(Gln->pyro-Glu)LSYGVDEK(+HLKNL)STGISVPGPM(Oxidation)GPSGPR	50.187	84.7	36.2	4.0e-19
1095.4994	4	1.5	Q(Gln->pyro-Glu)LSYGVDEK(+HLKNL)STGISVPGPM(Oxidation)GPSGPR	50.183	72.7	30.0	6.9e-17
1095.4987	4	0.88	Q(Gln->pyro-Glu)FDAK(+HLKNL)GGGPGPM(Oxidation)GLM(Oxidation)GPR	49.800	63.7	27.8	4.7e-16



MS-Viewer can now display results from results containing multiple different cross-linkers.

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 browser 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

¹Baker PR and Chalkley RJ, MS-Viewer: A Web Based Spectral Viewer for Proteomics Results. Molecular and Cellular Proteomics, (2014) doi:10.1074/mcp.O113.037200

Acknowledgements

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