

MS-Viewer: A Web-based Spectral Viewer for Proteomics Results*

Peter R. Baker‡§ and Robert J. Chalkley‡

The sharing and viewing of peptide identification results from search engines analyzing mass-spectrometry-based proteomic data is made difficult by the range of analysis tools employed, in that each produces a different output format. Annotated results associated with a journal article often have to be made available, but providing these in a format that can be queried by other researchers is often difficult. This is because although standard formats for results have been developed, these are not necessarily easy to produce. In this manuscript we describe the MS-Viewer program, part of the Protein Prospector Web package, which uses easy-to-create tabular files as input for providing highly interactive viewing of search engine results. Thanks to the simplicity and flexibility of the input format, results from a wide variety of search engines have been successfully viewed through the Web interface of this tool. *Molecular & Cellular Proteomics* 13: 10.1074/mcp.O113.037200, 1392–1396, 2014.

Mass spectrometric proteomic data are analyzed by a wide range of search engines. These software programs attach measures of reliability to results, and estimates of reliability at the dataset level are the most common threshold employed for the acceptance of results (1). However, if the search parameters employed were not appropriate (e.g. if the protein database queried did not contain many of the correct answers) or if the researcher focused the study on only a subset of the results (e.g. only post-translationally modified peptide identifications, which might not have the same reliability as the dataset as a whole), then the reliability statistics might be misleading (2). A biological researcher is generally most interested in one or a few results, so the ability to assess these before embarking on follow-up experiments is valuable. Typical results that may need more careful assessment are proteins identified by a single peptide and verifying post-translational modification identifications and site assignments, especially if

software was not employed to assess the reliability of modification site localizations independent of the measure of peptide identification reliability calculated by the search engine (3).

Publication guidelines for many proteomic journals require that annotated spectra be made available for some or all results (4), but as a plethora of different search engines are employed by researchers, each producing results files in a different format (5), there is not a simple mechanism to make results accessible and viewable by other researchers. Free spectral viewers can be downloaded and installed for results from a few analysis tools such as Scaffold (6) and OMSSA (7), and some labs have developed tools for displaying results from their search engines of choice (8). An alternative is to convert the output into a standard format. mzIdentML is a community-formulated standard format (9), but it is not widely employed yet. The repository PRIDE employs its own XML format (PRIDE XML), and tools have been written to convert outputs from several search engines into this format (10). These XML formats are designed to capture all details about parameters and results, so they are somewhat large and complicated to produce. Converters to standard XML formats are currently available for about half the search engines in common use.

One can make annotated spectra accessible, whether for supporting a publication or simply for sharing results with other researchers, using simple-to-create tabular files. In this manuscript we describe a spectral viewer (MS-Viewer) available through the Protein Prospector website. It allows annotated spectra from database search results to be viewed interactively. As the viewer reads tabular results files as input, it is minimal work to support the annotation of results from practically any search engine, making this a tool that can be used for viewing most peptide identification results. In addition, because it is hosted through a public website, it does not require the installation of any software (although a researcher may install a local version of the software for free if desired).

EXPERIMENTAL PROCEDURES

MS-Viewer requires two types of input files: a peak list file that contains the spectrum information, and a results file that contains the peptide assignments to each spectrum. In some XML results formats, such as PRIDE XML and pepxml, both of these information types are stored in the same file. The results should be in a single file, but if the dataset corresponds to multiple instrument runs, then multiple peak list files can be uploaded together in any common archive file format. All of the common peak list formats are supported: mgf, mzData, pkl, dta, mzML, mzXML, and ms2. Results files are uploaded in a tab-

From the ‡Mass Spectrometry Facility, Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94158

Received December 19, 2013, and in revised form, February 10, 2014

Published, MCP Papers in Press, March 3, 2014, DOI 10.1074/mcp.O113.037200

Author contributions: P.R.B. and R.J.C. contributed new reagents or analytic tools; P.R.B. and R.J.C. wrote the paper.

delimited text or comma-separated value format. Among the columns in this table, there must be one containing peptide sequences (with modifications either within the sequence or as a separate column), one with spectrum identifiers that allow mapping between the results file and the uploaded peak list file(s) (the identifier must be present in the results and peak lists in exactly the same format (e.g. if retention times in the peak list files are reported to three decimal places, then times in the results file must also have the same accuracy)), and one containing the precursor charge, which is used to determine which charge states are considered when annotating the spectrum. A fraction column (containing the name of the relevant peak list file) is also required if multiple peak lists are uploaded (again, for mapping between peak lists and results). An arbitrary number of other columns containing any other information may also be present.

These required columns are present in most results files from search engines. However, re-formatting of the content in these columns into a format understandable to MS-Viewer may be required. MS-Viewer expects peptide sequences to be in uppercase and modifications on peptides listed in the peptide sequence column to be described in round parentheses immediately after the modified residue. The only variation on this is that lowercase s, t, y, and m are interpreted as phosphorylation of ser, thr, and tyr or met oxidation, respectively. The modification itself can be either expressed using the PSI-MOD standard nomenclature (11) or reported as a mass (if a mass is reported, then this should be exact rather than nominal); for example, methionine oxidation can be indicated as M(Oxidation) or M(15.995). A modification on the N or C terminus is represented by a hyphen before or after the beginning or end residue; for example, an acetylated protein N-terminal peptide could be listed as Acetyl-MDESTR. The modification can also be described in a different column using the format "modification@residue_number_in_peptide_sequence." Using this format it is also possible to represent ambiguous modification site localizations, which can then be displayed and compared in MS-Viewer (as described later). Potential ambiguous site localizations should be separated by a vertical line; for example, if one wanted to represent that a phosphorylation could be on either the sixth or the seventh residue in a peptide sequence assignment, then this would be indicated as "Phospho@67." This format also supports neutral loss modifications (where the precursor is modified, but it is assumed that all fragments are unmodified); for example, Sulfo@Neutral loss would be the most effective way to annotate a sulfated peptide spectrum.

Most search engines can produce a tabular output, and as MS-Viewer cares about the format of only four of the columns, any required conversion can be achieved using a simple script or using spreadsheet software such as Microsoft Excel. Scripts to convert results outputs into a compliant format can either be run prior to upload or be automatically run by MS-Viewer; for example, scripts to convert Mascot csv, X!Tandem tab-delimited, and MSF (from the Thermo ProteomeDiscoverer software) output formats are implemented on the public website, so these can be uploaded without any further processing. Scripts to produce correctly formatted files from pepXML and PRIDE XML prior to submission to MS-Viewer have been implemented, so files in these formats can also be submitted directly to MS-Viewer, and results from these formats are included among the example datasets on the MS-Viewer webpage (<http://prospector.ucsf.edu/prospector/html/misc/viewereg.htm>). MS-Viewer can read BiblioSpec and NIST msp format files (12), so it can support viewing of spectral libraries. We are also aware that a Sequest user has converted results files into a format that was then displayed using MS-Viewer. Conversion scripts will be made available as links from the "help" section of MS-Viewer, and if researchers create their own converters, links to these will be provided as well.

Fig. 1 shows a screenshot of the upload page for MS-Viewer. More detailed instructions for uploading data to MS-Viewer are in the user

manual linked to from the website, and there is also a video tutorial demonstrating uploading from different formats (<http://vimeo.com/channels/194363>). It is also possible to run MS-Viewer via a command line interface, which may be a preferential option if the user has either large files to process or a large number of files requiring similar parameters. Details on how to do this are also included in the online instruction manual. Once files are uploaded and the submitter has set the default parameters to use for displaying results (e.g. fragment types to consider, fragment mass tolerance), then upon indication by the submitter that the results should be saved, a randomly generated search key is created. Only with knowledge of this search key can other people view the dataset. Thus, MS-Viewer is suitable for use during a journal manuscript review process, as the search key can be supplied to the reviewers and be made public by inclusion in the published manuscript.

As supplementary results files accompanying a manuscript are often spreadsheet files (i.e. delimited text files), an author could potentially use the same file for submission to the journal and MS-Viewer (13). It is also possible to create hyperlinks from a table outside of MS-Viewer directly to spectra (e.g. by pasting the HTML results table into Excel). An example of the use of outside links is DegraBase (14). However, browsing results through the MS-Viewer results interface is likely to be a more robust model for older datasets, as they are less vulnerable to future changes in software.

RESULTS

The results are browsed in a tabular format. As a default, the results are paginated with a defined number of rows per page. This is to reduce the page file size (viewing all rows for large datasets makes page loading slow, and browsers may crash). The ordering of results is according to how they were uploaded, but results can be re-sorted according to the content of any of the columns. It is also possible to filter results based on column content (either through the presence of a string or by displaying only columns with content greater or less than a certain numerical value). The user can also hide columns, so information of interest can more easily fit in the width of the browser window. MS-Viewer uses MS-Product to display annotated spectra. This is the software that is used to display results from MS/MS spectra by the Protein Prospector package, and it is unusually powerful and flexible. The following are some of the nonstandard features of this tool:

(a) Ability to compare different assignments to the same spectrum (it can compare up to six assignments on the same spectrum). This can be used to display the best and second best sequence assignments for comparison, or it can be used for viewing the ambiguity of modification site localization, if this information is supplied in the uploaded results file as described above. This process is completely interactive, so the user can change the assignment and get the software to re-annotate using the new sequence. It automatically displays ambiguous site localization assignments from Protein Prospector searches using SLIP scoring (15).

(b) Ability to threshold a peak list based on the number of peaks per half of the mass range or per m/z 100 bins. Some search engines use one of these approaches to threshold a peak list when scoring results, so the user can simulate

MS-Viewer

[Example MS-Viewer Datasets](#)

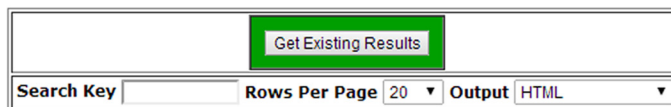
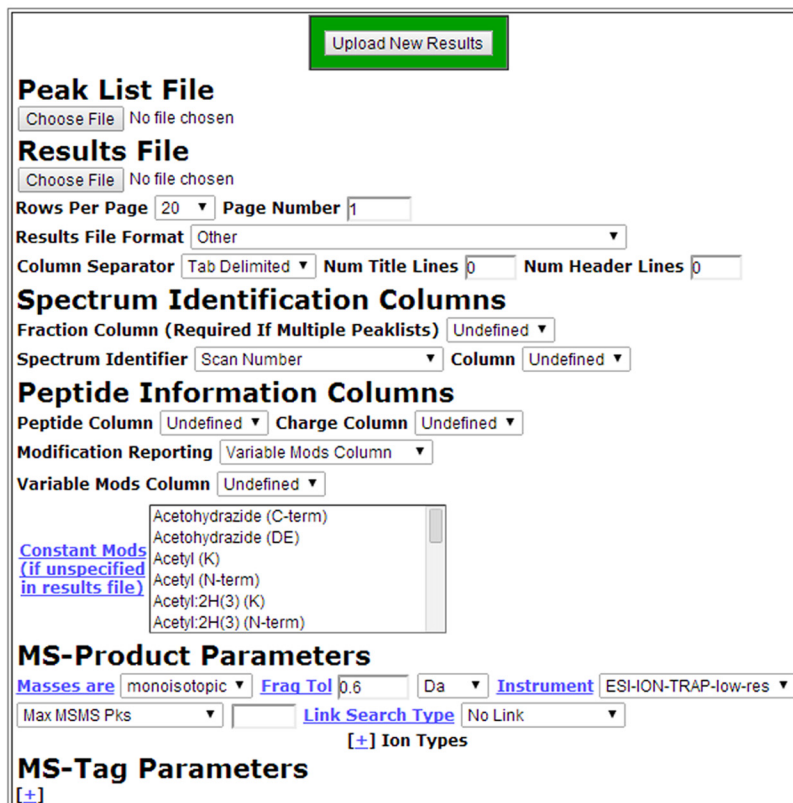


FIG. 1. Screenshot of the MS-Viewer submission page. At the top of the page is the link to previously submitted results, which can be viewed by inputting a search key for a particular dataset. Underneath are the fields for submitting new data. Users browse to select a peak list and results file. If they specify their results file format as Protein Prospector tab-delimited, Mascot csv, or X! Tandem tab-delimited text, then MS-Viewer is able to automatically recognize the columns it needs, whereas if 'Other' is selected the submitter must indicate in which columns required information is located. At the bottom of the page, users specify default parameters used for displaying and re-searching spectra.



what the search engine considered when making its peptide assignment.

(c) Ability to change fragment ion types considered. For example, MS-Viewer can be used to view ion trap collision-induced dissociation, quadrupole collision-induced dissociation, high-energy collision-induced dissociation, or electron-transfer/electron-capture dissociation data. In addition to sequence ions from single peptide backbone cleavages, it can label internal ions, immonium ions, and multiple neutral loss peaks (e.g. $y\text{-H}_2\text{O-NH}_3$).

(d) Ability to change the permitted fragment mass tolerance (in parts per million or daltons). The default tolerance is defined when the data are uploaded, but the user can change this value when viewing individual spectra.

(e) The user can recalibrate the spectrum based on the sequence fragment identifications. This can be useful when trying to differentiate between nominally isobaric alternatives (e.g. Q versus K, or trimethyl versus acetyl).

(f) Clicking on the spectrum identifier column in MS-Viewer allows the user to re-search individual spectra using MS-Tag in Protein Prospector, employing search parameters of his or her choice (e.g. the user could search against a different database or allow for different modifications), giving the researcher a second opinion on the reliability of the reported identification.

(g) MS-Viewer allows annotation of cross-linked peptide spectra (we believe no other viewer currently supports cross-linked peptide results). Each cross-linked peptide in the complex is reported in a separate column in the uploaded tabular file, where the mass of the second peptide with cross-linker is reported as a mass modification on the relevant residue in each peptide (13).

Fig. 2 shows an example of the spectral display in MS-Viewer for a HexNAc (O-GlcNAc) modified peptide where there was site localization ambiguity. The viewer indicates the evidence for the different site localizations. In this example, z6

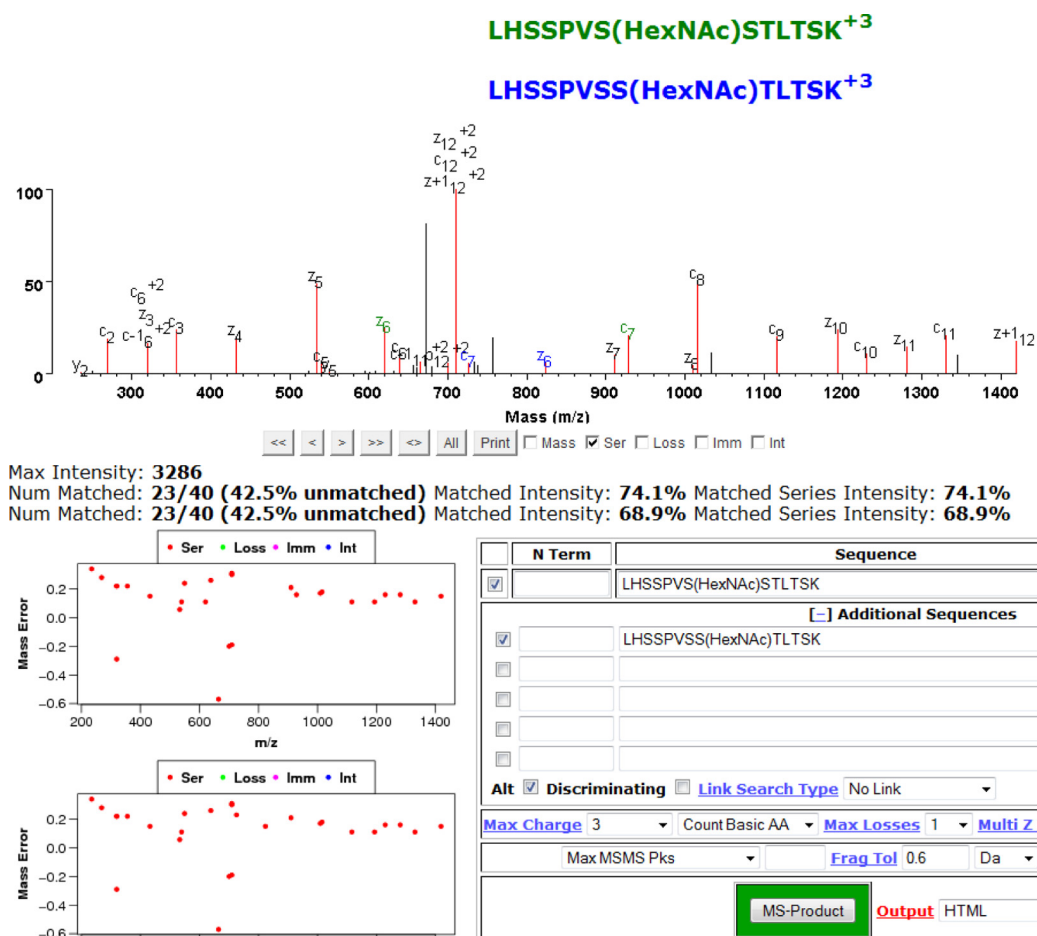


FIG. 2. Example of MS-Viewer displaying an annotated spectrum for a peptide in which the modification site localization was ambiguous. The peptide has two potential HexNAc modification site localizations shown in different colors; those peak assignments common to both are labeled in black, and those assignments unique to one or the other are labeled in the relevant color.

and c7 ions were observed for both compared site localizations, suggesting that peptide versions where each site was modified were probably co-fragmented.

DISCUSSION

The requirement for access to annotated spectra for results supporting journal publications has led to increased submission of results to public repositories, most notably to ProteomeXchange (5). ProteomeXchange currently receives two types of peptide identification submissions: those that are in the PRIDE XML format, allowing viewing of annotated spectra using PRIDE Inspector (16), and those labeled as “partial submissions,” in which the search engine results are not in a format that PRIDE can read. Many of these results formats can be displayed in MS-Viewer with little work; links to the results for several ProteomeXchange submissions that each used a different results format have been made available among the example MS-Viewer datasets, as proof of principle. Thus, MS-Viewer allows researchers to view results from datasets that may have no easy display option, although the results files are available. MS-Viewer will not necessarily be

able to support all ProteomeXchange submissions, as there is no guarantee that the minimal information required in order to display results will be present in a partial submission. Because the software and the data are stored on the same server and accessed through a Web browser, there is no need to install any software or download files (which may be very large) in order to view results, greatly reducing the burden associated with reviewing results.

A requirement for use with journal publications is that data cannot be manipulated or deleted after publication. Data are initially uploaded to a temporary folder. However, once a submitter saves the submission (needed to create a search key), he or she is no longer able to edit the content or displaying of the results and cannot delete content. We do not anticipate the deletion of any files on the public website; peak list and delimited text files are small in comparison to raw data files, and they also compress efficiently, so they can be stored with a smaller footprint and decompressed only when accessed. MS-Viewer could be installed at other locations, in which case data maintenance would be out of our control. However, possible other locations of installation could include

data repositories such as massIVE, which have extensive storage capabilities.

MS-Viewer can be accessed through the University of California San Francisco Protein Prospector website <http://prospector2.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msviewer>. Example datasets are also listed on that site <http://prospector.ucsf.edu/prospector/html/misc/viewereg.htm>. A video tutorial is available on Vimeo (<http://vimeo.com/30462677>).

* This work was supported by Grant No. 8P41GM103481 from NIGMS, National Institutes of Health.

§ To whom correspondence should be addressed: Peter R. Baker, 600 16th St., Genentech Hall, N474A, Box 2240, University of California San Francisco, San Francisco, CA 94158-2517. Tel.: 1-415-476-4893; Fax: 1-415-502-1655; E-mail: pbaker@cgl.ucsf.edu.

REFERENCES

1. Kall, L., Storey, J. D., MacCoss, M. J., and Noble, W. S. (2008) Assigning significance to peptides identified by tandem mass spectrometry using decoy databases. *J Proteome Res.* **7**, 29–34
2. Chalkley, R. J. (2013) When target-decoy false discovery rate estimations are inaccurate and how to spot instances. *J. Proteome Res.* **12**, 1062–1064
3. Chalkley, R. J., and Clauser, K. R. (2012) Modification site localization scoring: strategies and performance. *Mol. Cell. Proteomics* **11**, 3–14
4. Bradshaw, R. A., Burlingame, A. L., Carr, S., and Aebersold, R. (2006) Reporting protein identification data: the next generation of guidelines. *Mol. Cell. Proteomics* **5**, 787–788
5. Deutsch, E. W. (2012) File formats commonly used in mass spectrometry proteomics. *Mol. Cell. Proteomics* **11**, 1612–1621
6. Searle, B. C. (2010) Scaffold: a bioinformatic tool for validating MS/MS-based proteomic studies. *Proteomics* **10**, 1265–1269
7. Geer, L. Y., Markey, S. P., Kowalak, J. A., Wagner, L., Xu, M., Maynard, D. M., Yang, X., Shi, W., and Bryant, S. H. (2004) Open mass spectrometry search algorithm. *J. Proteome Res.* **3**, 958–964
8. Webber, J. T., Askenazi, M., and Marto, J. A. (2011) mzResults: an interactive viewer for interrogation and distribution of proteomics results. *Mol. Cell. Proteomics* **10**, M110.003970
9. Jones, A. R., Eisenacher, M., Mayer, G., Kohlbacher, O., Siepen, J., Hubbard, S. J., Selley, J. N., Searle, B. C., Shofstahl, J., Seymour, S. L., Julian, R., Binz, P. A., Deutsch, E. W., Hermjakob, H., Reisinger, F., Griss, J., Vizcaino, J. A., Chambers, M., Pizarro, A., and Creasy, D. (2012) The mzIdentML data standard for mass spectrometry-based proteomics results. *Mol. Cell. Proteomics* **11**, M111.014381
10. Vizcaino, J. A., Cote, R., Reisinger, F., Foster, J. M., Mueller, M., Rameseder, J., Hermjakob, H., and Martens, L. (2009) A guide to the Proteomics Identifications Database proteomics data repository. *Proteomics* **9**, 4276–4283
11. Montecchi-Palazzi, L., Beavis, R., Binz, P. A., Chalkley, R. J., Cottrell, J., Creasy, D., Shofstahl, J., Seymour, S. L., and Garavelli, J. S. (2008) The PSI-MOD community standard for representation of protein modification data. *Nat. Biotechnol.* **26**, 864–866
12. Frewen, B., and MacCoss, M. J. (2007) Using BiblioSpec for creating and searching tandem MS peptide libraries. *Curr. Protoc. Bioinformatics* Chapter 13, Unit 13.7
13. Trnka, M. J., Baker, P. R., Robinson, P. J., Burlingame, A. L., and Chalkley, R. J. (2014) Matching cross-linked peptide spectra: only as good as the worse identification. *Mol. Cell. Proteomics* **13**, 420–434
14. Crawford, E. D., Seaman, J. E., Agard, N., Hsu, G. W., Julien, O., Mahrus, S., Nguyen, H., Shimbo, K., Yoshihara, H. A., Zhuang, M., Chalkley, R. J., and Wells, J. A. (2013) The DegraBase: a database of proteolysis in healthy and apoptotic human cells. *Mol. Cell. Proteomics* **12**, 813–824
15. Baker, P. R., Trinidad, J. C., and Chalkley, R. J. (2011) Modification site localization scoring integrated into a search engine. *Mol. Cell. Proteomics* **10**, M111.008078
16. Wang, R., Fabregat, A., Rios, D., Ovelleiro, D., Foster, J. M., Cote, R. G., Griss, J., Csordas, A., Perez-Riverol, Y., Reisinger, F., Hermjakob, H., Martens, L., and Vizcaino, J. A. (2012) PRIDE Inspector: a tool to visualize and validate MS proteomics data. *Nat. Biotechnol.* **30**, 135–137