

11-1-2016

An Open Data Format for Visualization and Analysis of Cross-Linked Mass Spectrometry Results.

Michael R Hoopmann

Institute for Systems Biology, 401 Terry Avenue North, Seattle, WA 98109, USA.

Luis Mendoza

Institute for Systems Biology, Seattle, Washington, 98109, USA.

Eric W Deutsch

Institute for Systems Biology, Seattle, WA, USA.

David Shteynberg

Institute for Systems Biology, Seattle, Washington, 98109, USA.

Robert L Moritz

Institute for Systems Biology, Seattle, Washington, 98109, USA.

Follow this and additional works at: <https://digitalcommons.psjhealth.org/publications>



Part of the [Genetics and Genomics Commons](#)

Recommended Citation

Hoopmann, Michael R; Mendoza, Luis; Deutsch, Eric W; Shteynberg, David; and Moritz, Robert L, "An Open Data Format for Visualization and Analysis of Cross-Linked Mass Spectrometry Results." (2016). *Journal Articles and Abstracts*. 758.

<https://digitalcommons.psjhealth.org/publications/758>

This Article is brought to you for free and open access by Providence St. Joseph Health Digital Commons. It has been accepted for inclusion in Journal Articles and Abstracts by an authorized administrator of Providence St. Joseph Health Digital Commons. For more information, please contact digitalcommons@providence.org.



HHS Public Access

Author manuscript

J Am Soc Mass Spectrom. Author manuscript; available in PMC 2017 November 01.

Published in final edited form as:

J Am Soc Mass Spectrom. 2016 November ; 27(11): 1728–1734. doi:10.1007/s13361-016-1435-8.

An open data format for visualization and analysis of cross-linked mass spectrometry results

Michael R. Hoopmann^{1,*}, Luis Mendoza¹, Eric W. Deutsch¹, David Shteynberg¹, and Robert L. Moritz¹

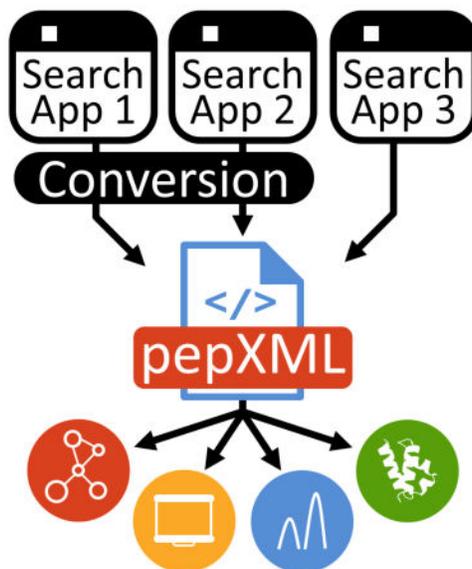
¹Institute for Systems Biology, Seattle, Washington 98109, USA

Abstract

Protein-protein interactions are an important element in the understanding of protein function, and chemical cross-linking shotgun mass spectrometry is rapidly becoming a routine approach to identify these specific interfaces and topographical interactions. Protein cross-link data analysis is aided by dozens of algorithm choices, but hindered by a lack of a common format for representing results. Consequently, interoperability between algorithms and pipelines utilizing chemical cross-linking remains a challenge. pepXML is an open, widely-used format for representing spectral search algorithm results that has facilitated information exchange and pipeline development for typical shotgun mass spectrometry analyses. We describe an extension of this format to incorporate cross-linking spectral search results. We demonstrate application of the extension by representing results of multiple cross-linking search algorithms. In addition, we demonstrate adapting existing pepXML-supporting software pipelines to analyze protein cross-linking results formatted in pepXML.

Graphical abstract

*Address correspondence to: Michael R. Hoopmann, Institute for Systems Biology, 401 Terry Ave N, Seattle, WA 98109, USA, michael.hoopmann@systemsbiology.org, Phone: 206-732-1200, Fax: 206-732-1299.



Introduction

Protein interactions and topologies are increasingly investigated with mass spectrometry (MS). Because of speed, simplicity, and affordability, chemical cross-linking followed by reversed-phase liquid chromatography and shotgun mass spectrometry (LC-MS/MS) has become one of the most common approaches for the analysis of protein interactions and three-dimensional structure [1–3]. In this approach, proteins and protein complexes are linked together using any of a myriad of available cross-linking reagents. The cross-linked proteins are then enzymatically digested to peptides and analyzed by LC-MS/MS, where cross-linked peptide sequences can be identified using specialized database search algorithms [4–5]. These cross-linked peptide spectrum matches (PSMs) contain two distinct peptides sequences that are used to infer proximal regions of protein structure from linearly distal domains, or domains from different proteins.

One of the biggest challenges when performing protein cross-linking and mass spectrometry (XL-MS) is data analysis. Parallel to the historical advances of shotgun MS, early efforts in cross-linked spectral analysis have given rise to a wealth of database search algorithms, too numerous to list fully, but include popular programs such as xQuest [6], pLink [7], Crux [8], Protein Prospector [9], Kojak [10], and StavroX [11]. Each software package has devised its own means of reporting results and, where applicable, visualizing them. Frequently, results and visualization tools for different algorithms are not readily transferrable, leaving further algorithm development and downstream analyses to proceed in isolation from the rest of the field.

A significant development in shotgun data analysis was the community desire to present results first in open common formats, such as pepXML [12], and then followed by open standardized formats, such as mzIdentML [13]. Through the utilization of common and standard formats, spectral search results from any algorithm could be easily contrasted,

visualized, and plugged into any analytical pipelines that make use of these common formats. Indeed, similar adoption of open formats (e.g. mzXML [14] and mzML [15]), allowed for analysis of data from multiple instruments and vendors because a common format was provided to represent the data. However, the existing formats for spectral search results do not currently allow for the representation of cross-linked PSMs, limiting the utility of results obtained from any particular cross-linking algorithm.

There are many examples of mass spectrometry software tools that become cornerstones for data analysis in many laboratories. Program suites such as the Trans-Proteomic Pipeline (TPP)[16] and Skyline [17], for example, have enjoyed widespread adoption in large part due to their free availability, open source architecture and adherence to open data standards and formats. Supporting these open formats makes the software accessible to a wider audience, promotes sharing and collaboration across laboratories, and drives community development of technology innovations. Thus, a critical step in the advancement of cross-linking analysis algorithms and pipelines is the establishment of a common format for passing information between software and laboratories.

Here we present an extension to the widely used, open data format, pepXML, for the storage and exchange of cross-linked PSMs. The extension supplements the current format, without impacting existing pepXML-based results from standard search algorithms. To illustrate the utility of the pepXML extension, we also present a new spectrum viewing application for cross-linked PSMs that, using the pepXML format, displays cross-linked search results from several different algorithms. Further use of pepXML is explored by demonstrating use of existing tools supporting pepXML to analyze cross-linking results, opening the possibility to incorporate multiple cross-linking algorithms into a common pipeline.

Materials and Methods

pepXML format extension

pepXML is an open file format for the storage of PSMs and subsequent peptide-level analysis [12]. The format makes use of the XML framework, which encapsulates elements of a shotgun MS/MS run with corresponding data analysis results. The schema for pepXML is provided at: <http://tools.proteomecenter.org/wiki/index.php?title=Formats:pepXML>. Within the format, *spectrum_query* elements contain MS/MS scan events, and encapsulate *search_result* elements that contain one or more *search_hit* elements relating to database search algorithm PSMs. The *search_hit* elements contain required information, such as peptide sequence, and further encapsulate other information such as static and differential modification masses and search algorithm scores.

To represent cross-linked PSMs in pepXML, multiple updates to the schema were required, shown in Figure 1 and listed in detail in Table 1. Because the *search_hit* element allows for only a single peptide sequence, a new attribute, *xlink_type*, was added. Valid values for the attribute are “xl” for two cross-linked peptides, “loop” for loop-linked (self-linked) PSMs, or “na” for single peptide PSMs. For backwards compatibility, the absence of the *xlink_type* attribute assumes a value of “na”. For “loop” and “na” values, all attributes pertaining to the PSM in the *search_hit* are used as normal. However, for “xl” values, these attributes are

ignored in favor of two *linked_peptide* elements contained inside the *search_hit* element, as described below. Two exceptions, the *calc_neutral_pep_mass* and *massdiff* attributes, describe the full cross-linked PSM mass and its difference in mass from the observed precursor ion, respectively (Supplementary Figure 1a).

When the *xlink_type* attribute is set to “loop” or “xl”, a novel *xlink* element must be provided inside the *search_hit*. The *xlink* must contain two attributes: an identifier for the cross-linker, and its mass contribution to the PSM. For self-linked peptides, *xlink_score* elements are used to describe the linked amino acid sites, and are contained within the *xlink* element (Supplementary Figure 1b). For cross-linked PSMs, two *linked_peptide* elements are used to describe both peptides. Many of the attributes of *linked_peptide* elements are the same as the *search_hit* element (i.e. peptide sequence, protein name, peptide neutral mass, etc.). Additional attributes include *complement_mass*, which is the difference in mass between the precursor ion and this peptide, and *designation*, which is used to label the peptide sequence as “alpha” or “beta”. The *linked_peptide* element encapsulates one or more *xlink_score* elements that are used to describe the linked amino acid site, plus any additional algorithm-specific scores that are only applicable to the individual peptides of the cross-linked PSM. Finally, any search algorithm scores applicable to the entire PSM, and not specifically to either cross-linked peptide, are specified in the typical *search_score* element contained within the *search_hit*.

Reference information for the cross-linker that is relevant to the search parameters or downstream analysis is also included in the pepXML extension. The new *cross_linker* element resides in the existing *msms_run_summary* element of pepXML. An *identifier* attribute names the cross-linker and should match the *identifier* attribute of *xlink* elements contained within *search_hit* elements. Additional attributes describe cross-linker mass, site reactivity, and isotopic labeling. The *cross_linker* element also contains a list of zero or more *cross_linker_info* elements for additional information specific to any particular cross-linker (e.g. spacer length) to be stored.

Analysis of MS/MS Spectra

LC-MS/MS data of human TFIH samples cross-linked with BS3 have been previously described and were obtained from the authors [18]. Briefly, samples were digested with trypsin, fractionated by strong cation exchange (SCX) and analyzed on a Thermo Fisher Scientific Orbitrap Velos with HCD fragmentation. The data were searched using multiple cross-linking database search algorithms: Kojak version 1.4.3 [10], pLink version 1.23 [7], and Protein Prospector version 5.16.0 [19]. Search parameters common to all algorithms included the 10 TFIH protein sequences, static and differential modification masses, search mass tolerances, and enzyme cleavage rules. Detailed listings of all parameters are provided in the Supplementary Information. Kojak results were exported to pepXML using the configuration parameter. Protein Prospector and pLink results were exported in tab-delimited text and converted to pepXML using conversion applications developed in house. Briefly, the conversion tools parse the tab-delimited text to identify precursor ion mass, precursor ion charge, peptide sequences, cross-linked sites, protein inferences, and PSM

score metrics. These values are then exported for all PSMs to the proposed pepXML format extension. The conversion applications are freely available at <http://www.kojak-ms.org/>.

Visualization of cross-linked PSMs

A graphical spectrum viewing application was designed to highlight matched fragment ions from cross-linked PSMs against acquired MS/MS spectra. The application accepts multiple data file formats (including mzXML and mzML) and PSMs from any search algorithm results formatted in pepXML. The software is written in C++, and is open source and freely available at <http://www.kojak-ms.org/>. For convenience, pre-compiled binary formats are also provided for both Windows and Linux operating systems. Though initially designed to be packaged and used with the Kojak cross-linking algorithm [10], the viewer is available in stand-alone format for use with results from any search algorithm.

Results and Discussion

The cross-linked data acquired from shotgun analysis of the human TFIID complex was analyzed using multiple database search algorithms as described in the methods. The set is comprised of 9 MS/MS runs totaling 92,348 spectra. Results from the Kojak algorithm were exported natively to the pepXML format using the configuration parameters. Results from pLink and Protein Prospector were converted from their default tab-delimited text formats to pepXML using the conversion tools described in the methods. Each search algorithm uses diverse score metrics to describe cross-linked PSMs. The pepXML format extension accounts for these differences, allowing the results from any of the algorithms to be represented in a common format. For example, pLink relies primarily on an expect score that is assigned to cross-linked PSMs, and the peptide-specific information is limited to differential modifications and site of linkage (Supplementary Figure 2). Protein Prospector and Kojak, on the other hand, have additional metrics applied to each cross-linked PSM and their component peptides (Supplementary Figures 3 and 4). Protein Prospector in particular provides both PSM-level algorithm scores and expect values, and additionally, peptide-level algorithm scores and expect values, whose metrics are useful in downstream validation processes. The pepXML schema extensions allow for all of this information to be stored in a common format, which can then be extracted as necessary by any downstream data analysis application.

Additionally, the algorithms pLink and Kojak provide PSMs for loop-linked and non-linked peptides, the latter often including peptides to which a cross-linker has bound but no linkage to another site was possible prior to quenching the reaction. Though not as useful as cross-linked PSMs, these identifications can be helpful when assessing solvent accessible sites on protein surfaces, or assessing experimental conditions. The pepXML schema extensions allow for these loop-linked and non-linked PSMs to be represented in the same results file, so that the entire data analysis from any search engine can be stored in a single file, and the different types of PSMs can be parsed as necessary using the attributes defined in the schema extension.

In accordance with the pepXML format, results from multiple searches can be combined into a single file, for example, when using the InteractParser from the TPP [12, 16]. The

TFIIH data were collected in multiple runs originating from SCX fractions prior to reversed phase LC-MS. pLink and Protein Prospector can combine their search results into a single output file, but Kojak cannot. After conversion of the results to pepXML, InteractParser was used to combine all the results into a single pepXML file for each search algorithm. Similarly, analyses which require multiple-pass searches to identify multiple cross-linker chemistries in the same data [20], can be combined into a single pepXML file because the format allows for multiple cross-linker and search parameter designations. Furthermore, pepXML files from multiple algorithm searches of the same data set can be combined into a single data set file, opening the possibility for multi-algorithm PSM validation analysis within the pepXML schema extension for cross-linked data sets. This technique has been used to improve peptide identification and validation in typical shotgun MS analyses [21].

Visualizing PSMs in pepXML files

To demonstrate the utility of a common format for the representation of cross-linked PSMs, and to facilitate visual inspection and evaluation of cross-linked PSMs, a spectrum visualization tool was developed. The visualization tool accepts as input a pepXML file. It then parses pepXML content to locate the corresponding data files and display the spectra in the view window. Overlaid on each spectrum are the fragment ions from the associated PSM search results (Figure 2). PSM information, such as peptide sequence, differential modifications, and fragment ions are provided in a customization pane on the right, and the user can toggle which information to highlight on top of the spectrum. Cross-linked PSMs containing two peptide sequences can be toggled to highlight the fragment ions from either or both peptides. Loop-linked PSMs identify both sites of cross-linker attachment and apply the appropriate mass corrections to the corresponding fragment ions. Non-linked PSMs, if provided in the search results, are also displayed, thus not limiting the spectrum visualization tool solely to cross-linked search results.

Inclusive in the visualization tool is a table summary of all PSMs contained in the pepXML. Each column in the table lists the scoring metrics of the respective algorithm that was used to generate the PSMs. Each row is a PSM. Selecting any row of the table instantly displays the spectrum and fragment ion masses of the corresponding peptide(s) in the PSM. The table can be customized to sort and filter the PSMs on the relevant score metrics for any algorithm. For example, top-ranked PSMs can be obtained by sorting expect values of pLink and Protein Prospector results, or the score value of Kojak results. Filters can be applied to limit the results to proteins of interest, or any particular metrics provided from the search results.

Compatibility with existing pipelines and alternative formats

Many tools currently exist that make use of the pepXML format for input and output, including PSM visualization, PSM validation, quantitation, and results organization [12, 17, 22–26]. By extending the pepXML format to include cross-linking search results, these tools become potentially extensible to analysis of cross-linked PSMs. Furthermore, it is far easier to update existing tools to include the pepXML extension, rather than recode them to support the many individual formats generated by each algorithm, or write a novel algorithm that replicates the same functionality for each of the cross-linking pipelines in existence. For

example, the TPP PepXML Viewer was upgraded to support this new cross-linking extension. The tool can now be used to browse, filter, categorize, and export custom spreadsheets for the cross-linking results from all cross-linking algorithms that make use of pepXML (Figure 3). In particular, the PepXML Viewer was used to create custom reports for the pepXML formatted results from the three algorithms used here. This is just one example of how a common pepXML format enables many different analysis tools to be plugged into cross-linking data analysis pipelines.

The proteomics field has made great strides in establishing open, community standards for data representation. The HUPO Proteomics Standards Initiative [27–28], has defined several open formats that allow data and results from many different instruments and pipelines to be represented in a common language. The current standard for database search results, mzIdentML, does not support cross-linked PSMs at the time of this writing. The pepXML schema extension presented here provides a means to support the increasingly common cross-linked peptide analyses, until such time that a community standard is established. The transition to new standards is also not immediate, as is seen with the persistence of other data formats such as MGF and mzXML, and the tools that require use of them. Thus, support for cross-linked PSMs in pepXML will help ease the burden of the eventual transition to mzIdentML by extending the life of existing tools and pipelines that support pepXML until upgrades and alternatives are developed.

Conclusions

pepXML is an open format in which results from many different search algorithms are represented uniformly, enabling easy integration of diverse results into robust analytical pipelines. The schema for pepXML was extended to represent results from shotgun cross-linking spectral searching algorithms. With these schema extensions, we demonstrate how results from different cross-linking search algorithms can be easily transformed for use by downstream software applications, such as the cross-linking spectral viewer. By extending an open format, it is possible to apply many existing tools and pipelines to cross-linking data analysis, as demonstrated with the minor upgrades to the TPP. The format extension presented here has the potential to increase integration of cross-linking analysis in mass spectrometry analytical workflows.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank Drs. Jie Luo and Jeff Ranish for access to the TFIID data analyzed in this publication. This work was funded in part by National Institutes of Health from the National Institute of General Medical Sciences under grant nos. R01GM087221, S10RR027584, and the 2P50 GM076547/Center for Systems Biology.

References

1. Bruce JE. In vivo protein complex topologies: sights through a cross-linking lens. *Proteomics*. 2012; 12:1565–1575. [PubMed: 22610688]

2. Sinz A. The advancement of chemical cross-linking and mass spectrometry for structural proteomics: from single proteins to protein interaction networks. *Expert Rev Proteomics*. 2014; 11:733–743. [PubMed: 25227871]
3. Walzthoeni T, Leitner A, Stengel F, Aebersold R. Mass spectrometry supported determination of protein complex structure. *Curr Opin Struct Biol*. 2013; 23:252–260. [PubMed: 23522702]
4. Mayne SL, Patterson HG. Bioinformatics tools for the structural elucidation of multi-subunit protein complexes by mass spectrometric analysis of protein-protein cross-links. *Brief Bioinform*. 2011; 12:660–671. [PubMed: 22101029]
5. Sinz A, Arlt C, Chorev D, Sharon M. Chemical cross-linking and native mass spectrometry: A fruitful combination for structural biology. *Protein Sci*. 2015; 24:1193–1209. [PubMed: 25970732]
6. Rinner O, Seebacher J, Walzthoeni T, Mueller LN, Beck M, Schmidt A, Mueller M, Aebersold R. Identification of cross-linked peptides from large sequence databases. *Nat Methods*. 2008; 5:315–318. [PubMed: 18327264]
7. Yang B, Wu YJ, Zhu M, Fan SB, Lin J, Zhang K, Li S, Chi H, Li YX, Chen HF, Luo SK, Ding YH, Wang LH, Hao Z, Xiu LY, Chen S, Ye K, He SM, Dong MQ. Identification of cross-linked peptides from complex samples. *Nat Methods*. 2012; 9:904–906. [PubMed: 22772728]
8. McIlwain S, Draghicescu P, Singh P, Goodlett DR, Noble WS. Detecting cross-linked peptides by searching against a database of cross-linked peptide pairs. *J Proteome Res*. 2010; 9:2488–2495. [PubMed: 20349954]
9. Trnka MJ, Baker PR, Robinson PJ, Burlingame AL, Chalkley RJ. Matching cross-linked peptide spectra: only as good as the worse identification. *Mol Cell Proteomics*. 2014; 13:420–434. [PubMed: 24335475]
10. Hoopmann MR, Zelter A, Johnson RS, Riffle M, MacCoss MJ, Davis TN, Moritz RL. Kojak: efficient analysis of chemically cross-linked protein complexes. *J Proteome Res*. 2015; 14:2190–2198. [PubMed: 25812159]
11. Gotze M, Pettelkau J, Schaks S, Bosse K, Ihling CH, Krauth F, Fritzsche R, Kuhn U, Sinz A. StavroX—a software for analyzing crosslinked products in protein interaction studies. *J Am Soc Mass Spectrom*. 2012; 23:76–87. [PubMed: 22038510]
12. Keller A, Eng J, Zhang N, Li XJ, Aebersold R. A uniform proteomics MS/MS analysis platform utilizing open XML file formats. *Mol Syst Biol*. 2005; 1:20050017.
13. Jones AR, Eisenacher M, Mayer G, Kohlbacher O, Siepen J, Hubbard SJ, Selley JN, Searle BC, Shofstahl J, Seymour SL, Julian R, Binz PA, Deutsch EW, Hermjakob H, Reisinger F, Griss J, Vizcaino JA, Chambers M, Pizarro A, Creasy D. The mzIdentML data standard for mass spectrometry-based proteomics results. *Mol Cell Proteomics*. 2012; 11:M111014381.
14. Pedrioli PG, Eng JK, Hubley R, Vogelzang M, Deutsch EW, Raught B, Pratt B, Nilsson E, Angeletti RH, Apweiler R, Cheung K, Costello CE, Hermjakob H, Huang S, Julian RK, Kapp E, McComb ME, Oliver SG, Omenn G, Paton NW, Simpson R, Smith R, Taylor CF, Zhu W, Aebersold R. A common open representation of mass spectrometry data and its application to proteomics research. *Nat Biotechnol*. 2004; 22:1459–1466. [PubMed: 15529173]
15. Martens L, Chambers M, Sturm M, Kessner D, Levander F, Shofstahl J, Tang WH, Rompp A, Neumann S, Pizarro AD, Montecchi-Palazzi L, Tasman N, Coleman M, Reisinger F, Souda P, Hermjakob H, Binz PA, Deutsch EW. mzML—a community standard for mass spectrometry data. *Mol Cell Proteomics*. 2011; 10:R110000133.
16. Deutsch EW, Mendoza L, Shteynberg D, Slagel J, Sun Z, Moritz RL. Trans-Proteomic Pipeline, a standardized data processing pipeline for large-scale reproducible proteomics informatics. *Proteomics Clin Appl*. 2015; 9:745–754. [PubMed: 25631240]
17. MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, Kern R, Tabb DL, Liebler DC, MacCoss MJ. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics*. 2010; 26:966–968. [PubMed: 20147306]
18. Luo J, Cimermancic P, Viswanath S, Ebmeier CC, Kim B, Dehecq M, Raman V, Greenberg CH, Pellarin R, Sali A, Taatjes DJ, Hahn S, Ranish J. Architecture of the Human and Yeast General Transcription and DNA Repair Factor TFIIH. *Mol Cell*. 2015; 59:794–806. [PubMed: 26340423]

19. Chalkley RJ, Baker PR, Medzihradszky KF, Lynn AJ, Burlingame AL. In-depth analysis of tandem mass spectrometry data from disparate instrument types. *Mol Cell Proteomics*. 2008; 7:2386–2398. [PubMed: 18653769]
20. Leitner A, Joachimiak LA, Unverdorben P, Walzthoeni T, Frydman J, Forster F, Aebersold R. Chemical cross-linking/mass spectrometry targeting acidic residues in proteins and protein complexes. *Proc Natl Acad Sci U S A*. 2014; 111:9455–9460. [PubMed: 24938783]
21. Shteynberg D, Deutsch EW, Lam H, Eng JK, Sun Z, Tasman N, Mendoza L, Moritz RL, Aebersold R, Nesvizhskii AI. iProphet: multi-level integrative analysis of shotgun proteomic data improves peptide and protein identification rates and error estimates. *Mol Cell Proteomics*. 2011; 10:M111007690.
22. Kall L, Canterbury JD, Weston J, Noble WS, MacCoss MJ. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat Methods*. 2007; 4:923–925. [PubMed: 17952086]
23. Ma ZQ, Dasari S, Chambers MC, Litton MD, Sobecki SM, Zimmerman LJ, Halvey PJ, Schilling B, Drake PM, Gibson BW, Tabb DL. IDPicker 2.0: Improved protein assembly with high discrimination peptide identification filtering. *J Proteome Res*. 2009; 8:3872–3881. [PubMed: 19522537]
24. Mohammed Y, Mostovenko E, Henneman AA, Marissen RJ, Deelder AM, Palmblad M. Cloud parallel processing of tandem mass spectrometry based proteomics data. *J Proteome Res*. 2012; 11:5101–5108. [PubMed: 22916831]
25. Park CY, Klammer AA, Kall L, MacCoss MJ, Noble WS. Rapid and accurate peptide identification from tandem mass spectra. *J Proteome Res*. 2008; 7:3022–3027. [PubMed: 18505281]
26. Park SK, Aslanian A, McClatchy DB, Han X, Shah H, Singh M, Rauniyar N, Moresco JJ, Pinto AF, Diedrich JK, Delahunty C, Yates JR 3rd. Census 2: isobaric labeling data analysis. *Bioinformatics*. 2014; 30:2208–2209. [PubMed: 24681903]
27. Orchard S, Hermjakob H, Apweiler R. The proteomics standards initiative. *Proteomics*. 2003; 3:1374–1376. [PubMed: 12872238]
28. Deutsch EW, Albar JP, Binz PA, Eisenacher M, Jones AR, Mayer G, Omenn GS, Orchard S, Vizcaino JA, Hermjakob H. Development of data representation standards by the human proteome organization proteomics standards initiative. *J Am Med Inform Assoc*. 2015; 22:495–506. [PubMed: 25726569]

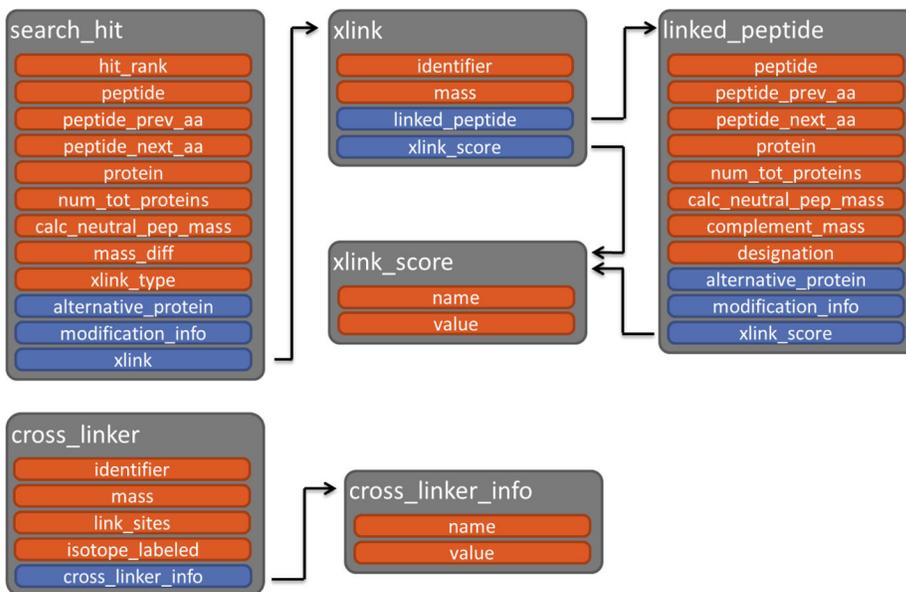


Figure 1. pepXML schema extensions to represent cross-linked search results. Elements are represented in gray. Within each element, attributes are colored in red and nested elements are colored in blue. The *search_hit* element was extended to include the *xlink_type* attribute and nested *xlink* element. The *cross_linker* element includes zero or more *cross_linker_info* elements to describe additional features specific to any particular cross-linker.

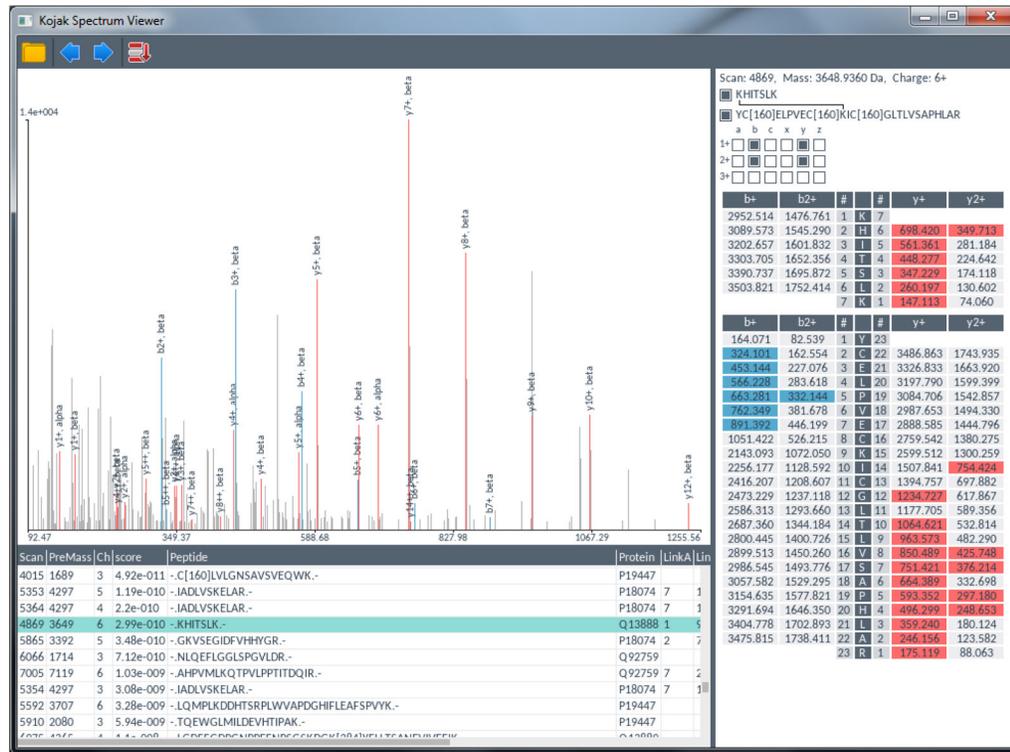


Figure 2. Image of the spectrum viewer. PSMs for any search algorithm are read from pepXML files. The PSMs are filtered, sorted, and selected using the table at the bottom left. Visual elements of the spectral match are toggled and highlighted using the display customization pane on the right. When displaying cross-linked PSMs, the user can toggle highlighting matched fragment ions of both peptides.

SPECTRUM	CALC_MASS	KLEW_TYPE	EXPECT	SCORE	SCORE_DEPTH	PEPTIDE	PROTEIN
VE122812_HumanIH_40Ac-11.4716.4716.4	4880.4145	xl	4.7e-22	101.3 53.0 42.5	47.7	[spectrum] - KAVVDFEAMIDNWClee.øIDSMSWLTB.-R - ILDAKQKGLVHP56TGR.-R	P18074 P18074
VE122812_HumanIH_40Ac-11.4727.4727.5	4880.4145	xl	4.8e-19	114.3 63.6 48.0	50.7	[spectrum] - KAVVDFEAMIDNWClee.øIDSMSWLTB.-R - ILDAKQKGLVHP56TGR.-R	P18074 P18074
VE122812_HumanIH_40Ac-11.3016.3016.4	2578.3576	xl	1.1e-16	92.7 21.1 67.8	24.9	[spectrum] - FSKVDGK.-R - KNLCLee.øTHPEVTPLR.-R	P18074 P18074
VE122812_HumanIH_40Ac-11.2995.2995.5	2578.3576	xl	6.8e-16	84.5 15.5 61.8	22.7	[spectrum] - FSKVDGK.-R - KNLCLee.øTHPEVTPLR.-R	P18074 P18074
VE122812_HumanIH_40Ac-11.5287.5287.4	4297.2084	xl	7.6e-16	81.7 6.3 66.1	7.3	[spectrum] - IADLVKELAR.-R - KAVVDFEAMIDNWClee.øIDSMSWLTB.-R	P18074 P18074
VE122812_HumanIH_40Ac-11.3874.3874.3	3088.6304	xl	9.6e-16	76.7 72.8 -1.3	3.9	[spectrum] - IALLGGKcse.17AHSDDTSQ6PDKcH8.-R - KB.-R	Q92759 Q92759
VE122812_HumanIH_40Ac-11.5495.5495.5	4137.1059	xl	1.7e-15	72.5 17.4 46.0	26.5	[spectrum] - AEKLTLSGNPS.-R - YCLee.øELPVECLee.øKTCLee.øSLTLVSAPHLAR.-R	Q13888 Q13888
VE122812_HumanIH_40Ac-11.3189.3189.4	3282.6699	xl	2.7e-15	78.5 18.9 49.1	29.4	[spectrum] - ISKFEVDMDEK.-R - SKLAVTASHIQESS.-R	Q13889 Q13889
VE122812_HumanIH_40Ac-11.3832.3832.5	3175.6624	xl	3.4e-15	85.2 80.5 -5.5	4.7	[spectrum] - IALLGGKcse.17AHSDDTSQ6PDKcH8.-R - KSR.-R	Q92759 Q92759

Figure 3. A screen capture of the PepXML Viewer tool in the TPP. The tool can display search results from any cross-linking algorithm that have been converted to pepXML. Tabs along the top of the viewer can be used to select metrics to display, sort, and filter, and the tool can be used to export custom summary reports of cross-linked search results.

Table 1

Attribute descriptions for cross-linking pepXML extensions

Element: search_hit	
Attribute	Description
xlink_type	values of “na”, “loop”, and “xl” to indicate PSM is non-linked, loop-linked, or cross-linked, respectively.
Element: xlink	
Attribute	Description
identifier	name of the cross-linker in ASCII text, e.g. “DSS” or “BS3”.
mass	mass contribution of the cross-linker.
Element: linked_peptide	
Attribute	Description
peptide	amino acid sequence of the peptide.
previde_prev_aa	amino acid preceding the peptide in this protein.
peptide_next_aa	amino acid following the peptide in this protein.
protein	protein name from the database.
num_tot_proteins	total number of proteins that share this peptide. Additional proteins are listed in the <i>alternative_protein</i> element already described in pepXML.
calc_neutral_pep_mass	uncharged mass of this peptide in the PSM.
complement_mass	remaining mass of the other peptide and cross-linker.
designation	“alpha” or “beta” for describing the first or second peptide of a cross-linked PSM.
Element: xlink_score	
Attribute	Description
name	text label for the score.
value	text or numerical value for the score.
Element: cross_linker	
Attribute	Description
identifier	name of the cross-linker in ASCII text, e.g. “DSS” or “BS3”.
mass	mass contribution of the cross-linker.
link_sites	amino acid(s) targeted by cross-linker, in upper case. May include multiple amino acids, e.g. “DE”. Lower-case “n” or “c” used to include protein termini, e.g. “nK”.
isotope_labeled	“n” for no, “y” for yes, to indicate if cross-linker has been heavy isotope labeled.
Element: cross_linker_info	
Attribute	Description
name	text label describing cross-linker feature, e.g. “spacer_arm_length”
value	text or numerical value for the feature