

ProXL (Protein Cross-Linking Database): A Platform for Analysis, Visualization, and Sharing of Protein Cross-Linking Mass Spectrometry Data

Michael Riffle,^{*,†,‡} Daniel Jaschob,[†] Alex Zelter,[†] and Trisha N. Davis[†]

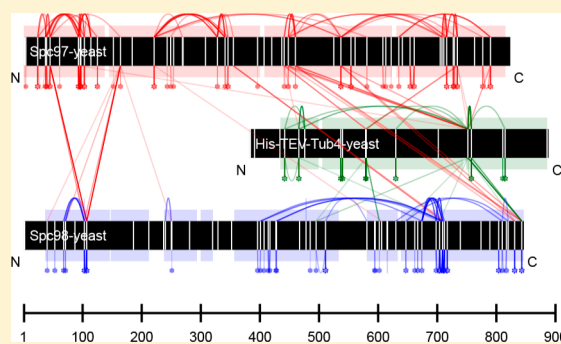
[†]Department of Biochemistry, University of Washington, Seattle, Washington 98195, United States

[‡]Department of Genome Sciences, University of Washington, Seattle, Washington 98195, United States

S Supporting Information

ABSTRACT: ProXL is a Web application and accompanying database designed for sharing, visualizing, and analyzing bottom-up protein cross-linking mass spectrometry data with an emphasis on structural analysis and quality control. ProXL is designed to be independent of any particular software pipeline. The import process is simplified by the use of the ProXL XML data format, which shields developers of data importers from the relative complexity of the relational database schema. The database and Web interfaces function equally well for any software pipeline and allow data from disparate pipelines to be merged and contrasted. ProXL includes robust public and private data sharing capabilities, including a project-based interface designed to ensure security and facilitate collaboration among multiple researchers. ProXL provides multiple interactive and highly dynamic data visualizations that facilitate structural-based analysis of the observed cross-links as well as quality control. ProXL is open-source, well-documented, and freely available at <https://github.com/yeastrc/proxl-web-app>.

KEYWORDS: cross-linking, data visualization, structure, proteomics, software, bioinformatics, database



INTRODUCTION

Understanding a protein's structure is fundamental to understanding that protein's function. Identifying interaction partners, sites of interaction, and the structural architecture of multiprotein complexes is fundamental to determining their role in cellular processes. Protein cross-linking coupled with bottom-up mass spectrometry (XL-MS or CX-MS or CLMS) has been gaining ground in recent years as a tool for elucidating the structure, architecture, and dynamics of large multiprotein complexes.^{1–11} XL-MS differs from traditional bottom-up tandem mass spectrometry (MS/MS) in that the protein mixture is subjected to a chemical cross-linker prior to digestion and analysis by mass spectrometry (for reviews, see refs 12–14). The chemical cross-linker binds protein residues on both ends and has a spacer arm of known length in between. After analysis and identification of linked peptides by mass spectrometry, the linked peptides and positions may be mapped to proteins and indicate which residues in those proteins are near one another in solution. These known distances within and between specific positions in proteins may serve as unique distance restraints (UDRs) when used in conjunction with molecular modeling, protein structure prediction, or other structure-based methods.^{9,15–17}

Although XL-MS promises a wealth of structural information, the automated identification of cross-linked peptides from tandem mass spectra has been a difficult computational

problem. The cross-linking reaction and digestion may produce different species of peptides (cross-linked, loop-linked, and unlinked) (Figure 1) that have different scoring characteristics. Moreover, the search space for candidate pairs of cross-linked peptides matching a precursor ion's m/z is prohibitively large for sequence databases typically used in proteomics analysis. Several algorithms have made significant progress toward addressing this complexity and have been widely adopted for automated XL-MS analysis, including Kojak,¹⁸ pLink,¹⁹ Crux,²⁰ xQuest,²¹ StavroX,²² Protein Prospector,²³ SIM-XL,²⁴ and Hekate.²⁵ While these software packages enable many researchers to identify cross-linked peptides and proteins, the visual interfaces to the data and results are limited. Because each software package produces its own proprietary scores and reads and writes its own file formats, the data are usually not portable or compatible with interfaces provided by other software. Direct comparison of results from separate software packages (or even different versions of the same software) is difficult.

Several visualization tools have been developed to extend the data visualization capabilities provided by the native XL-MS search software. Xlink Analyzer²⁶ is a software extension to the UCSF Chimera²⁷ molecular modeling software package that

Received: April 1, 2016

Published: June 15, 2016

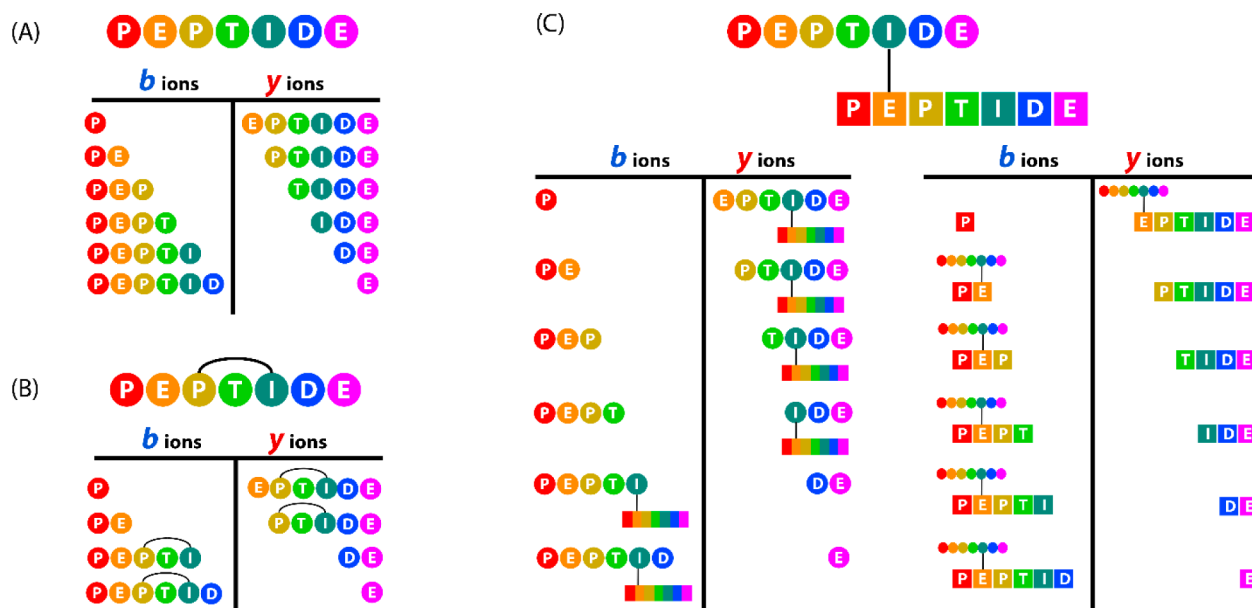


Figure 1. Depiction of the b- and y-ion series generated from bottom-up XL-MS peptide fragmentation and supported by ProXL. ProXL treats monolinks (where only one end of a cross-linker has reacted with a peptide residue) as a special case of a post-translational modification (PTM). Monolinked residues may be found in all three peptide species. (A) An “unlinked” peptide. The peptide has no residues linked to any other residues. The ion series is typical of those found in non-XL-MS experiments. (B) A “loop-linked” peptide. The single peptide contains two residues linked to one another. The ion series does not include breaks between the two linked residues. (C) “Cross-linked” peptides. A residue in one peptide is linked to a residue in another peptide. There are separate b- and y-ion series for each of the linked peptides, where the mass of the other peptide is considered as a modification of the mass of the linked residue, as if it were a PTM.

enables import, visualization, and structural analysis of reported UDRs. xiNET²⁸ is a dynamic web application and Javascript library that provides dynamic and compelling two-dimensional views of XL-MS search results. It ingeniously combines the traditional network topology display of protein–protein interactions found in tools like Cytoscape²⁹ with scaled horizontal bars representing the lengths of individual proteins found in protein sequence annotation tools. xVis³⁰ is a web application that provides elegant two-dimensional network topology visualization of XL-MS results, including a topology display similar to xiNET and CIRCOS-style³¹ displays of the data. Unlike Xlink Analyzer, xVis does not depend on third-party software for visualization, and unlike both Xlink Analyzer and xiNET, xVis provides direct access to the underlying proteomics data (e.g., mass spectra). However, this functionality depends on xQuest for data analysis and the availability of a local xQuest server. XLink-DB³² is a web application and database for storing, viewing, and disseminating XL-MS results. It includes two-dimensional (2D) and three-dimensional (3D) visualization and analysis tools. While its emphasis on public dissemination and visualization of data from any pipeline is a step in the right direction, it depends on third-party plugins to function, depends on the use of the UniProt³³ database for protein sequence annotation, and is (at the time of this writing) limited to experiments from *Escherichia coli*, *Homo sapiens*, *Saccharomyces cerevisiae*, and *Arabidopsis thaliana*.

Here we present ProXL, a web application and database for storing, visualizing, and sharing XL-MS data that is cross-platform and independent of search software and protein sequence annotation database, does not require third-party software, provides integrated access to all underlying proteomics data, and functions equally well for any organism. ProXL provides dynamic 2D and 3D visualization, reporting, and analysis tools, including quality control tools and data

downloads for optional integration into third-party software for more advanced analysis. ProXL includes advanced data sharing tools, both public and private, and is designed to enable collaboration among project researchers.

DESIGN AND IMPLEMENTATION

Technology

ProXL consists of a web application, relational database, and data import program. The web application was developed using Java, HTML, CSS, SVG, and Javascript and was designed to run on the Apache Tomcat (<http://tomcat.apache.org/>) Java servlet container and the Struts application framework (<http://struts.apache.org/>). The built-in Protein Data Bank (PDB) structure viewer uses the pv structure viewer (<https://github.com/biasmv/pv>),³⁴ which is pure Javascript, and requires no third-party plugins to run. Spectra are visualized by a version of the Lorikeet spectrum viewer (<http://uwpr.github.io/Lorikeet/>)³⁵ that we have modified to view loop-linked and cross-linked ion series. Real-time protein sequence annotations for disordered regions and secondary structure prediction are provided by Disopred3³⁶ and Psipred3³⁷ and are executed in response to user requests by the JobCenter job management system³⁸ running on the authors' servers. The relational database was developed using the MySQL (<https://www.mysql.com/>) relational database management system. The import program was developed using Java and XML (see below for more information on the XML schema). All of the components of ProXL are cross-platform and will run on any platform for which Java is available.

Installation

On the client side, there is no installation required for users of ProXL (other than using a current web browser). All viewers and functionality are written using standard Worldwide Web

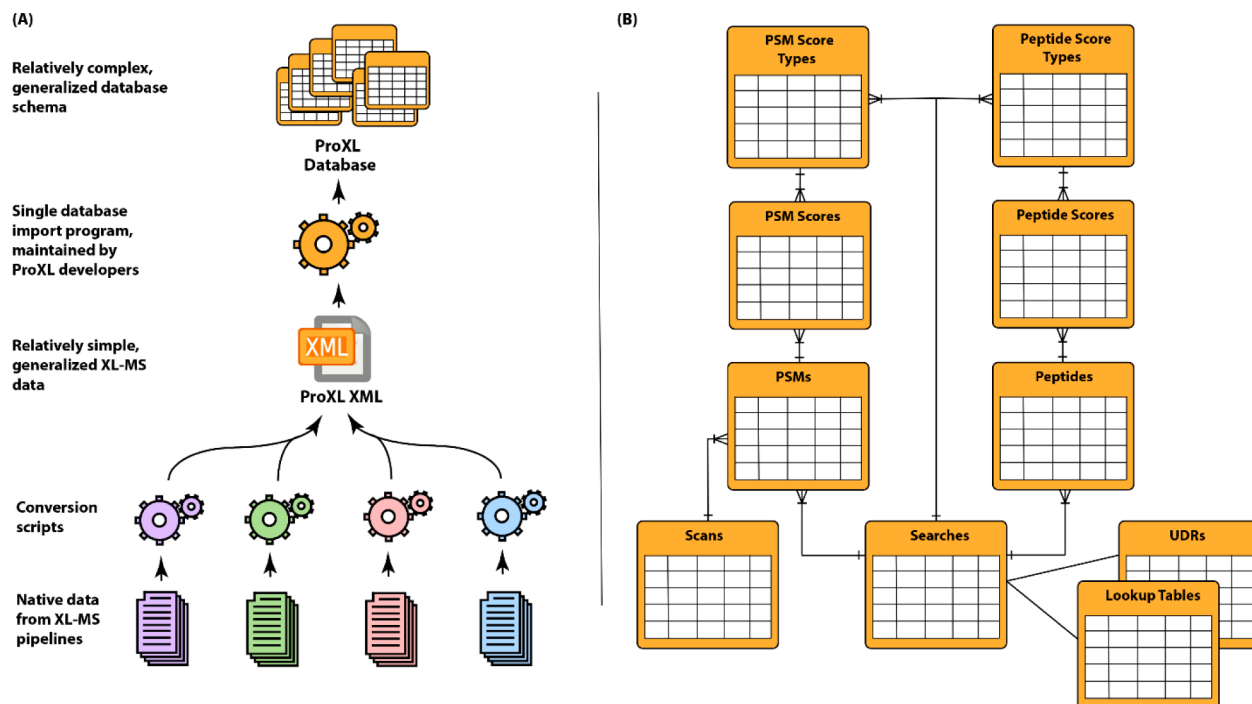


Figure 2. Overview of the ProXL data flow and database design. (A) Different software pipelines produce data files using their own disparate formats. Writing and maintaining programs to import this native data directly into the complex ProXL database schema would be complex, error-prone, and difficult for developers of new pipelines. Instead, simple scripts are used to convert native data into the simple and well-documented ProXL XML format, which is a generalized format for representing XL-MS data and includes descriptions of which scores are present and how the scores are to be treated by ProXL. A central program, maintained by the ProXL developers, is used to import ProXL XML files into the database. (B) This cartoon overview of the database schema illustrates how ProXL generalizes the association of scores from any pipeline with XL-MS data. Score types described by ProXL XML files are stored in score-type tables, where the names, descriptions, and properties of scores present in the XML are stored. Scores for PSMs and peptides are associated with both this score type and a generalized abstraction of PSMs and peptides applicable to all search programs (e.g., sequence for peptides). PSMs are associated with scans, PSMs and peptides with searches, and searches with UDR information, lookup tables, and other generalized attributes. See [Figure S-1](#) for the true database schema.

technologies and do not require any external programs or web browser plugins. On the server side, ProXL makes use of multiple database and web application components and will require basic knowledge of MySQL and system administration to install and configure. Specifically, Apache Tomcat and MySQL will need to be installed (if not already installed), SQL scripts executed to set up the database, and values changed in the database to configure the web application. Full installation instructions are available at the ProXL documentation website (http://yeastrc.org/proxl_docs/).

Data Design and Import

ProXL's data design is independent of any particular software pipeline that generates cross-link search results ([Figure 2](#)). This is accomplished by abstracting types of data common to all cross-linking pipelines into a common set of core tables that describe items such as the identified UDRs (i.e., which protein loci were found to be linked to one another), the scan data, and the identified peptide sequences. The scores assigned to peptide spectrum matches (PSMs) and peptides from individual software pipelines are stored in score tables that describe which search program was used, which scoring attributes were present, how to treat those scoring attributes, and what score for which attribute each PSM or peptide received. A diagram of the ProXL database schema is shown in [Figure S-1](#) in the [Supporting Information](#).

Importing data into the ProXL database is accomplished by converting the native output of the respective software pipelines into an XML file adhering to the ProXL XML

schema ([Figure S-2](#)). Like the ProXL database schema, this XML schema is independent of any particular software pipeline. Which PSM-level and peptide-level scoring attributes are present for the respective software pipeline are included and described in the XML, including how to label it, how to sort it, and default filter values. This design allows the output of nearly any conceivable pipeline, regardless of the type of numeric scores generated, to be represented in this XML format and imported into ProXL for visualization, analysis, and comparison. Additionally, using this XML schema as a common standard for importing data dramatically simplifies the process of developing and maintaining importers for new software pipelines, as developers are shielded from the complexity of the database schema itself. The schema includes XML schema validation rules designed to help ensure the integrity of the data in the file. The schema XML schema definition (XSD) file, documentation, and programs for converting Kojak (with and without the Percolator³⁹ postsearch analysis), Crux, Plink, StavroX, and XQuest output to ProXL XML are available at <https://github.com/yeastrc/proxl-import-api>.

ProXL is also designed to be independent of the specific FASTA file or sequence database used to search peptides. The FASTA data files are preprocessed before the data is uploaded to ProXL so that the strings representing proteins are mapped back to a nonredundant protein sequence database. The advantages are twofold. Any sequence database can be used to generate the FASTA data files. Proteins identified in different experiments using different search databases data can be

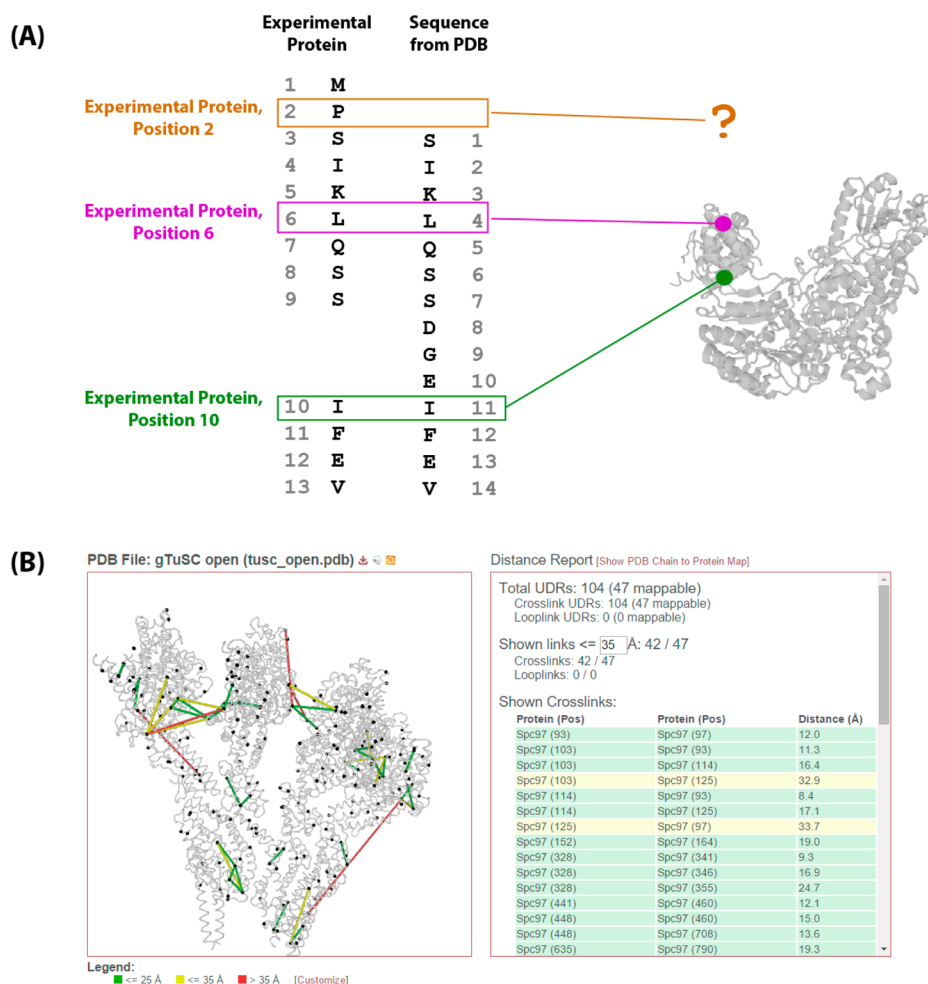


Figure 3. Structure alignment and display in ProXL. (A) ProXL allows users to map proteins identified in the experiment to the sequences present in any PDB-formatted file. This is accomplished by an automated (and user-validated) pairwise sequence alignment, which may result in gaps or shifts in the respective alignments. Positions in the experimental protein use this alignment to map to positions in the PDB sequence, which are then used to map the position to 3D space. Positions in the experimental protein that do not map to the PDB sequence are considered “unmappable” and are not represented on the structure or used for distance reports. (B) Screenshot from ProXL illustrating a mapping of three *S. cerevisiae* proteins (Spc97p, Spc98p, and Tub4p) to a dimer of the yeast small γ -tubulin complex. Two copies each of Spc97p and Spc98p and four copies of Tub4p are present in the structure. The left panel shows the 3D structure and may be zoomed or rotated. The links are color-coded according to calculated distances. The right panel shows a distance report for the currently displayed data, which is color-coded to match the links on the structure. All of the links (in either panel) may be clicked to view underlying proteomics data or spectra; the PDB file, UCSF Chimera script, or PyMOL script may be downloaded, and distance reports may be downloaded via links at the bottom of the report.

directly compared. We have developed a web application to ease the preprocessing of FASTA files. Usage and installation documentation for this application are available at our documentation site (http://yeastrc.org/proxl_docs/).

ProXL Data Visualization Tools

ProXL includes HTML tables and dynamic, graphical views of the data. In all cases, data are linked to the underlying proteomics data, including annotated spectra. For example, a table of identified UDRs in a given run shows which positions in which proteins were found to be linked to one another. If additional information about the identification is required, the row may be expanded to view all of the underlying peptides. This may be further expanded to view all of the underlying PSMs and associated spectra. For all graphical views, the current state of the viewer (i.e., all selected options and protein positions) is encoded in the current URL for the web page. Because of the breadth of options and complexity of the data, significant time may be invested in achieving the desired view of

the data. Whenever an option changes, the URL is automatically updated to reflect the change in state. As such, this URL may be bookmarked or shared with other users who have access to the project to simplify collaboration and sharing of specific views of the data. Additionally, for all views, the current view of the data (including all options) may be saved as the default view of a given search's data for the viewer, allowing the desired view of the data to be shared with other users.

Summaries of various views of the data provided by ProXL are outlined below. Detailed documentation of all features is available at the ProXL documentation site (http://yeastrc.org/proxl_docs/) or by clicking the help icon near the top right of any page in ProXL.

3D Structure View. ProXL's 3D structure viewer allows cross-links, loop-links, and monolinks to be visualized on interactive graphical representations of protein structures (Figure 3b). This is accomplished by providing tools for users to upload a PDB file (whether their own or from the PDB) and then automatically perform pairwise sequence

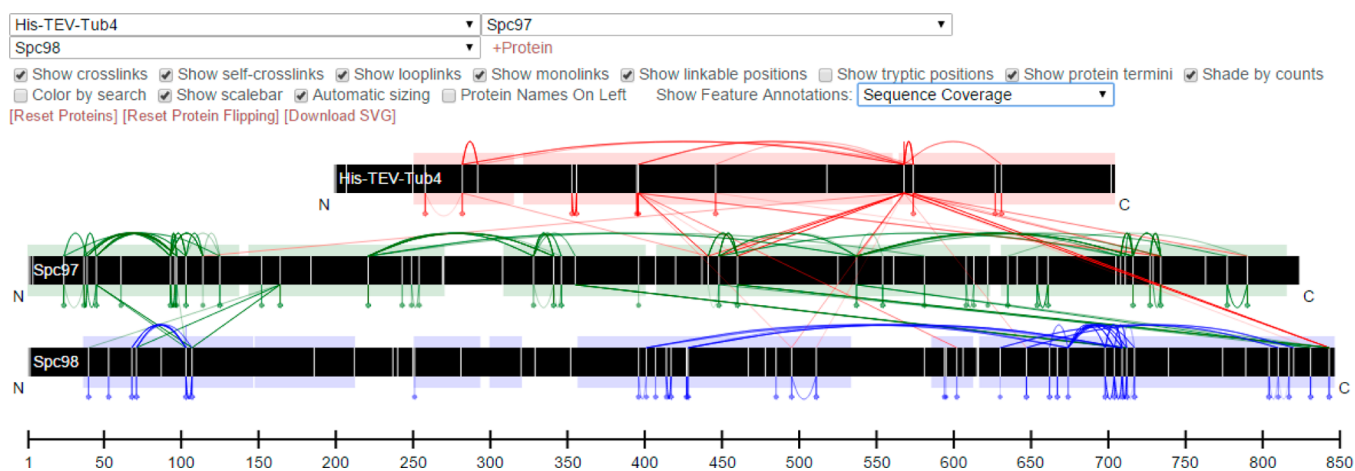


Figure 4. Screen shot from ProXL of the “image view” display of linked horizontal protein bars representing protein sequences. The black bars, from top to bottom, represent respective sequence lengths for Tub4p, Spc97p, and Spc98p from the *S. cerevisiae* small γ -tubulin complex. The lines between bars represent interprotein cross-links, the arcs above the bars represent intraprotein cross-links, the arcs beneath the bars represent loop-links, and the balls and sticks on the bottom represent monolinks. The shaded regions on the protein bars represent sequence coverage. The white vertical lines on the bars represent sites that may react with the cross-linker used in the experiment. The colors are unique to each protein, with cross-links between proteins being colored according to the protein above it in the diagram. The user may click on all of the links to view underlying proteomics data and spectra. The interface includes many customization options, which are fully documented at http://yeastrc.org/proxl_docs/.

alignment between protein sequences from their search’s FASTA file to sequences present for chains in the PDB. This alignment is used to map identified link locations for proteins in the FASTA file to specific locations in the PDB structure, enabling 3D visualization and distance measurements of observed links (Figure 3a). This design does not require that the exact sequence from a PDB chain be present in the FASTA file used to search the data and allows for the use of PDB files containing single proteins or multiprotein complexes—where all of the proteins in the complex may be mapped to proteins from the experiment.

This structure may be rotated, zoomed, and recentered. All visualized links may be clicked to view the underlying peptide- and PSM-level data, including annotated mass spectra. Distances for all represented links are calculated and available for viewing as a table or downloaded as a text file. The PDB file and link locations may be downloaded as PyMOL⁴⁰ or UCSF Chimera scripts for visualization and analysis in the respective software.

Features of note include (1) the ability to pop the structure out into a separate window for high-resolution viewing and figure generation, (2) the ability to shade the displayed links by spectrum counts, and (3) the ability to color the observed links on the basis of their calculated distance, their type (cross-link or loop-link), or (in the case of merging of multiple searches) the search(es) in which that link was observed at the current cutoff values. This enables quick side-by-side structure-based comparison of multiple searches (including different search programs). All of the observed links and distances are also displayed as a table, and these data may be downloaded as a tab-delimited text report.

This view is available in the “Explore Data” section of the project page by clicking the “[Structure]” link associated with a given search. Multiple searches can be combined by selecting multiple searches and clicking the “View Merged Structure” button.

Graphical Protein Bars. ProXL provides an interactive, customizable, and dynamic 2D view of the data. Proteins are displayed as horizontal bars scaled to their relative lengths

(Figure 4). Proteins found in the experiment may be added or removed from the display by the user. Interprotein cross-links are presented as line segments connecting the linked locations in each protein. Intraprotein cross-links and loop-links are presented as loops on the top and bottom of the protein bars, respectively, and monolinks are presented as short line segments. By default, links are colored according to protein to ease interpretation. When data from multiple searches are merged, the links may be colored according to the originating search to ease comparison of the searches. The links may be shaded according to spectrum counts to perform basic relative quantitative estimations. To aid in the interpretation of complex diagrams, a single protein may be clicked to highlight only links involving that protein, or multiple proteins may be clicked to highlight only the links on and between those proteins.

The protein bars may be moved horizontally, rescaled, and flipped. In addition, protein bars may be annotated by sequence coverage, predicted disordered regions, and predicted secondary structure—with the latter two options being run in real time for proteins without these annotations in the database.

This view is available in the “Explore Data” section of the project page by clicking the “[Image]” link associated with a given search. Multiple searches can be combined by selecting multiple searches and clicking the “View Merged Image” button.

Quality Control. Quality control for cross-linking proteomics experiments is complex because cross-linked and loop-linked peptides are evaluated in addition to the unlinked peptides found in traditional proteomics experiments. Differences in experimental design, mass spectrometry performance, or search software may affect the behavior and identification of cross-linked, loop-linked, or unlinked peptides differently. ProXL provides two quality control visualizations for assessing the relative performance of these different classes of peptides (Figure S-3). One visualization assesses the performance of peptide identifications as a function of retention time. Total scans and the number of scans resulting in quality identifications are plotted versus retention time. The user

may select the score to be used for analysis and the cutoff value for the “quality score”. The other visualization assesses the performance of peptide identifications as a function of PSM quality scores (e.g., q-value or XCorr) by showing the cumulative total of identified PSMs as a function of score. The user may select which score from the experiment is used.

These visualizations are available in the ProXL interface in the “Explore Data” section of the project page by expanding a given search and clicking either the “[Retention Time]” or “[PSM Scores]” links next to “QC Plots.”

Tabular Data and Downloads. In addition to the visual display of data presented above, ProXL provides the data in table form, including all observed cross-links and loop-links (UDRs), all identified peptides, and all PSMs. In all cases, rows in tables may be expanded to view the supporting proteomics data and scores from the search. For example, rows in the cross-link table may be expanded to view all of the identified peptides and scores from the search that indicated that UDR. These peptides may themselves be expanded to view all of the underlying PSMs and scores for each peptide, and the spectrum associated with each PSM may be viewed. As with the graphical views, multiple searches may be combined and compared in table form (Figure S-4). The specific search(es) that identified the respective UDR, peptide, or PSM are indicated, and all levels of data (protein, peptide, and PSM) are clearly differentiated by search to simplify comparison. Additionally, all levels of data may be downloaded as tab-delimited text for use in other types of analysis, such as modeling.

These visualizations are available in the ProXL interface in the “Explore Data” section of the project page by clicking either the “[Peptides]” or “[Proteins]” link associated with a given run.

Data Sharing and Collaboration

ProXL organizes access to data by projects (Figure S-5). A project may be created by any user of ProXL, and a title, an abstract, users, and data can then be associated with that project. To associate researchers with the project, users may refer directly to existing users or supply e-mail addresses for new users. Existing users are immediately added to the project, and new users are invited by e-mail and may use a link to register and access the project. Researchers may leave notes or comments about the data that are visible to other researchers on the project.

Most critically, a project serves to limit access to the data. By default, public access is disabled and access to the data is limited to those users associated with that project. This ensures that only researchers associated with the given collaboration may access the data. Data may be optionally shared with researchers who do not have ProXL accounts by enabling public access on the project. The most restrictive form of public access requires that external users use a specially formatted URL containing an unguessable key that provides access to the project and its data. This ensures that only individuals who have been given this URL may access the data. The least restrictive form of public access does not require the unguessable key in the URL, making the URL much shorter and more appropriate for referencing in articles. Public access may be enabled and disabled by the project owner through the project overview page in the ProXL web interface.

Because each project has a unique URL, public project pages may be used as landing pages for sharing data associated with published articles. To facilitate this, projects may be locked by

the project owner, which prohibits any further changes to that project, including uploading data, changing public access levels, or altering the title or abstract.

USE AND IMPACT

ProXL has been in production use with numerous collaborators, who have driven its development and helped identify and resolve issues. As of this writing, the authors’ installation of ProXL contains 24 cross-linking projects comprising 135 mass spectrometry runs (searches). These searches found 8 099 216 PSMs from 6 800 113 distinct scans, identifying 1 385 093 distinct UDRs from cross-links and 22 862 distinct UDRs from loop-links.

An example of how ProXL may be used and of its impact is illustrated by the study by Zelter et al.,⁹ in which cross-linking mass spectrometry was combined with computational structural modeling to determine the molecular architecture of the *S. cerevisiae* Dam1 kinetochore complex. ProXL’s visualization tools, search comparison, and data download tools were critical tools for the authors to evaluate the quality of cross-linking experiments. It was used to visualize the differences of observed cross-links across different experimental conditions and to export data for use by external analysis tools and the Integrative Modeling Platform (IMP),⁴¹ the software platform used by the authors to predict the structure of the Dam1 complex from the cross-linking data. Finally, ProXL’s data sharing tools were used to publically disseminate the data (including RAW, post-processed, and cross-linking visualization) as a companion to the published article. The public ProXL site for that paper may be found at <http://proxl.yeastrc.org/dam1-zelter-2015>.

FUTURE DIRECTIONS

ProXL is actively used and developed. New features are regularly added and are driven by the needs of collaborators and users. Features and directions currently under development include new visual displays (including dynamic network topologies and CIRCOS-style views), support for other structural formats (including NMR or output from 3D modeling platforms), and quantification tools. We expect that the ProXL XML format will be extended to support new types of cross-linking data (e.g., quantification data), and we hope to work directly with the community to develop ProXL XML conversion tools for more software platforms.

CONCLUSIONS

ProXL is a web application and database designed to store, visualize, compare, and share cross-linking mass spectrometry data. ProXL is independent of any software analysis pipeline or FASTA sequence naming database. It has been designed to simplify the development of import tools for new pipelines and includes tools to combine and compare data generated from disparate pipelines. ProXL provides visualization tools particularly suited to structural analysis and quality control, tools for exporting the data, and data sharing tools designed for both private collaboration and public data dissemination. For demonstration purposes, a public ProXL project has been set up at http://yeastrc.org/proxl_demo/. ProXL is thoroughly documented, open-source, and freely available at <https://github.com/yeastrc/proxl-web-app/>.

■ ASSOCIATED CONTENT

■ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jproteome.6b00274.

Entity relationship diagram of the ProXL database schema (Figure S-1); diagram of the ProXL XML schema (Figure S-2); screenshots of the quality control plots from ProXL (Figure S-3); screenshot from the ProXL “Merged Proteins” page (Figure S-4); screenshot from the ProXL project page (Figure S-5) (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*Address: UW Box 357350, 1705 NE Pacific Street, Seattle, WA 98195-7350. Phone: 206-685-3740. E-mail: mriffle@uw.edu.

Notes

The authors declare no competing financial interest.

Author e-mail addresses: djaschob@uw.edu (D.J.); azelter@uw.edu (A.Z.); tdavis@uw.edu (T.N.D.).

■ ACKNOWLEDGMENTS

This work was supported by the National Institute of General Medical Sciences of the National Institutes of Health (P41 GM103533) and the University of Washington Proteomics Resource (UWPR95794).

■ ABBREVIATIONS:

MS/MS, tandem mass spectrometry; XML, Extensible Markup Language; XSD, XML schema definition; SQL, Structured Query Language; JSP, Java server page; HTML, Hypertext Markup Language; CSS, cascading style sheet; SVG, scalable vector graphics; PSM, peptide spectrum match; PDB, Protein Data Bank; UDR, unique distance restraint; XL-MS, protein cross-linking mass spectrometry; m/z , mass-to-charge ratio

■ REFERENCES

- (1) Tomko, R. J., Jr.; Taylor, D. W.; Chen, Z. A.; Wang, H. W.; Rappasilber, J.; Hochstrasser, M. A Single alpha Helix Drives Extensive Remodeling of the Proteasome Lid and Completion of Regulatory Particle Assembly. *Cell* **2015**, *163* (2), 432–44.
- (2) Luo, J.; Cimermancic, P.; Viswanath, S.; Ebmeier, C. C.; Kim, B.; Dehecq, M.; Raman, V.; Greenberg, C. H.; Pellarin, R.; Sali, A.; Taatjes, D. J.; Hahn, S.; Ranish, J. Architecture of the Human and Yeast General Transcription and DNA Repair Factor TFIIH. *Mol. Cell* **2015**, *59* (5), 794–806.
- (3) Dodonova, S. O.; Diestelkoetter-Bachert, P.; von Appen, A.; Hagen, W. J.; Beck, R.; Beck, M.; Wieland, F.; Briggs, J. A. VESICULAR TRANSPORT. A structure of the COPI coat and the role of coat proteins in membrane vesicle assembly. *Science* **2015**, *349* (6244), 195–8.
- (4) Greber, B. J.; Bieri, P.; Leibundgut, M.; Leitner, A.; Aebersold, R.; Boehringer, D.; Ban, N. Ribosome. The complete structure of the 55S mammalian mitochondrial ribosome. *Science* **2015**, *348* (6232), 303–8.
- (5) Greber, B. J.; Boehringer, D.; Leibundgut, M.; Bieri, P.; Leitner, A.; Schmitz, N.; Aebersold, R.; Ban, N. The complete structure of the large subunit of the mammalian mitochondrial ribosome. *Nature* **2014**, *515* (7526), 283–6.
- (6) Erzberger, J. P.; Stengel, F.; Pellarin, R.; Zhang, S.; Schaefer, T.; Aylett, C. H.; Cimermancic, P.; Boehringer, D.; Sali, A.; Aebersold, R.

Ban, N. Molecular architecture of the 40S eIF1eIF3 translation initiation complex. *Cell* **2014**, *158* (5), 1123–35.

- (7) Murakami, K.; Elmlund, H.; Kalisman, N.; Bushnell, D. A.; Adams, C. M.; Azubel, M.; Elmlund, D.; Levi-Kalishman, Y.; Liu, X.; Gibbons, B. J.; Levitt, M.; Kornberg, R. D. Architecture of an RNA polymerase II transcription pre-initiation complex. *Science* **2013**, *342* (6159), 1238724.

- (8) Herzog, F.; Kahraman, A.; Boehringer, D.; Mak, R.; Bracher, A.; Walzthoeni, T.; Leitner, A.; Beck, M.; Hartl, F. U.; Ban, N.; Malmstrom, L.; Aebersold, R. Structural probing of a protein phosphatase 2A network by chemical cross-linking and mass spectrometry. *Science* **2012**, *337* (6100), 1348–52.

- (9) Zelter, A.; Bonomi, M.; Kim, J. O.; Umbreit, N. T.; Hoopmann, M. R.; Johnson, R.; Riffle, M.; Jaschob, D.; MacCoss, M. J.; Moritz, R. L.; Davis, T. N. The molecular architecture of the Dam1 kinetochore complex is defined by cross-linking based structural modelling. *Nat. Commun.* **2015**, *6*, 8673.

- (10) Tien, J. F.; Umbreit, N. T.; Zelter, A.; Riffle, M.; Hoopmann, M. R.; Johnson, R. S.; Fonslow, B. R.; Yates, J. R., 3rd; MacCoss, M. J.; Moritz, R. L.; Asbury, C. L.; Davis, T. N. Kinetochore biorientation in *Saccharomyces cerevisiae* requires a tightly folded conformation of the Ndc80 complex. *Genetics* **2014**, *198* (4), 1483–93.

- (11) Kudalkar, E. M.; Scarborough, E. A.; Umbreit, N. T.; Zelter, A.; Gestaut, D. R.; Riffle, M.; Johnson, R. S.; MacCoss, M. J.; Asbury, C. L.; Davis, T. N. Regulation of outer kinetochore Ndc80 complex-based microtubule attachments by the central kinetochore Mis12/MIND complex. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (41), E5583–9.

- (12) Liu, F.; Heck, A. J. Interrogating the architecture of protein assemblies and protein interaction networks by cross-linking mass spectrometry. *Curr. Opin. Struct. Biol.* **2015**, *35*, 100–108.

- (13) Holding, A. N. XL-MS: Protein cross-linking coupled with mass spectrometry. *Methods* **2015**, *89*, 54–63.

- (14) 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* (8), 1193–209.

- (15) Webb, B.; Lasker, K.; Velazquez-Muriel, J.; Schneidman-Duhovny, D.; Pellarin, R.; Bonomi, M.; Greenberg, C.; Raveh, B.; Tjioe, E.; Russel, D.; Sali, A. Modeling of proteins and their assemblies with the Integrative Modeling Platform. *Methods Mol. Biol.* **2014**, *1091*, 277–95.

- (16) Shi, Y.; Pellarin, R.; Fridy, P. C.; Fernandez-Martinez, J.; Thompson, M. K.; Li, Y.; Wang, Q. J.; Sali, A.; Rout, M. P.; Chait, B. T. A strategy for dissecting the architectures of native macromolecular assemblies. *Nat. Methods* **2015**, *12* (12), 1135–8.

- (17) Politis, A.; Stengel, F.; Hall, Z.; Hernandez, H.; Leitner, A.; Walzthoeni, T.; Robinson, C. V.; Aebersold, R. A mass spectrometry-based hybrid method for structural modeling of protein complexes. *Nat. Methods* **2014**, *11* (4), 403–6.

- (18) Hoopmann, M. R.; Zelter, A.; Johnson, R. S.; Riffle, M.; MacCoss, M. J.; Davis, T. N.; Moritz, R. L. Kojak: efficient analysis of chemically cross-linked protein complexes. *J. Proteome Res.* **2015**, *14* (5), 2190–8.

- (19) Yang, B.; Wu, Y. J.; Zhu, M.; Fan, S. B.; Lin, J.; Zhang, K.; Li, S.; Chi, H.; Li, Y. X.; Chen, H. F.; Luo, S. K.; Ding, Y. H.; Wang, L. H.; Hao, Z.; Xiu, L. Y.; Chen, S.; Ye, K.; He, S. M.; Dong, M. Q. Identification of cross-linked peptides from complex samples. *Nat. Methods* **2012**, *9* (9), 904–6.

- (20) McIlwain, S.; Draghicescu, P.; Singh, P.; Goodlett, D. R.; Noble, W. S. Detecting cross-linked peptides by searching against a database of cross-linked peptide pairs. *J. Proteome Res.* **2010**, *9* (5), 2488–95.

- (21) Rinner, O.; Seebacher, J.; Walzthoeni, T.; Mueller, L. N.; Beck, M.; Schmidt, A.; Mueller, M.; Aebersold, R. Identification of cross-linked peptides from large sequence databases. *Nat. Methods* **2008**, *5* (4), 315–8.

- (22) Gotze, M.; Pettelkau, J.; Schaks, S.; Bosse, K.; Ihling, C. H.; 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* (1), 76–87.

(23) Chalkley, R. J.; Baker, P. R.; Medzihradzky, K. F.; Lynn, A. J.; Burlingame, A. L. In-depth analysis of tandem mass spectrometry data from disparate instrument types. *Mol. Cell. Proteomics* **2008**, *7* (12), 2386–98.

(24) Lima, D. B.; de Lima, T. B.; Balbuena, T. S.; Neves-Ferreira, A. G.; Barbosa, V. C.; Gozzo, F. C.; Carvalho, P. C. SIM-XL: A powerful and user-friendly tool for peptide cross-linking analysis. *J. Proteomics* **2015**, *129*, 51–5.

(25) Holding, A. N.; Lamers, M. H.; Stephens, E.; Skehel, J. M. Hekate: software suite for the mass spectrometric analysis and three-dimensional visualization of cross-linked protein samples. *J. Proteome Res.* **2013**, *12* (12), 5923–33.

(26) Kosinski, J.; von Appen, A.; Ori, A.; Karius, K.; Muller, C. W.; Beck, M. Xlink Analyzer: software for analysis and visualization of cross-linking data in the context of three-dimensional structures. *J. Struct. Biol.* **2015**, *189* (3), 177–83.

(27) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25* (13), 1605–12.

(28) Combe, C. W.; Fischer, L.; Rappsilber, J. xiNET: cross-link network maps with residue resolution. *Mol. Cell. Proteomics* **2015**, *14* (4), 1137–47.

(29) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13* (11), 2498–504.

(30) Grimm, M.; Zimniak, T.; Kahraman, A.; Herzog, F. xVis: a web server for the schematic visualization and interpretation of crosslink-derived spatial restraints. *Nucleic Acids Res.* **2015**, *43* (W1), W362–9.

(31) Krzywinski, M.; Schein, J.; Birol, I.; Connors, J.; Gascoyne, R.; Horsman, D.; Jones, S. J.; Marra, M. A. Circos: an information aesthetic for comparative genomics. *Genome Res.* **2009**, *19* (9), 1639–45.

(32) Zheng, C.; Weisbrod, C. R.; Chavez, J. D.; Eng, J. K.; Sharma, V.; Wu, X.; Bruce, J. E. XLink-DB: database and software tools for storing and visualizing protein interaction topology data. *J. Proteome Res.* **2013**, *12* (4), 1989–95.

(33) The UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Res.* **2015**, *43* (D1), D204–12.

(34) Biasini, M. PV - JavaScript Protein Viewer. <https://biasini.github.io/pv/> (accessed April 1, 2016).

(35) Sharma, V.; Eng, J. K.; Maccoss, M. J.; Riffle, M. A mass spectrometry proteomics data management platform. *Mol. Cell. Proteomics* **2012**, *11* (9), 824–31.

(36) Jones, D. T.; Cozzetto, D. DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinformatics* **2015**, *31* (6), 857–63.

(37) Jones, D. T. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **1999**, *292* (2), 195–202.

(38) Jaschob, D.; Riffle, M. JobCenter: an open source, cross-platform, and distributed job queue management system optimized for scalability and versatility. *Source Code Biol. Med.* **2012**, *7* (1), 8.

(39) Kall, L.; Canterbury, J. D.; Weston, J.; Noble, W. S.; MacCoss, M. J. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat. Methods* **2007**, *4* (11), 923–5.

(40) *The PyMOL Molecular Graphics System*, version 1.8; Schrödinger, LLC: New York, 2015.

(41) Russel, D.; Lasker, K.; Webb, B.; Velazquez-Muriel, J.; Tjioe, E.; Schneidman-Duhovny, D.; Peterson, B.; Sali, A. Putting the pieces together: integrative modeling platform software for structure determination of macromolecular assemblies. *PLoS Biol.* **2012**, *10* (1), e1001244.