PyXlinkViewer: A flexible tool for visualization of protein chemical crosslinking data within the PyMOL molecular graphics system

Bob Schiffrin | Sheena E. Radford | David J. Brockwell | Antonio N. Calabrese

Abstract
Chemical crosslinking-mass spectrometry (XL-MS) is a valuable technique for gaining insights into protein structure and the organization of macromolecular complexes. XL-MS data yield inter-residue restraints that can be compared with high-resolution structural data. Distances greater than the crosslinker spacer-arm can reveal lowly populated “excited” states of proteins/protein assemblies, or crosslinks can be used as restraints to generate structural models in the absence of structural data. Despite increasing uptake of XL-MS, there are few tools to enable rapid and facile mapping of XL-MS data onto high-resolution structures or structural models. PyXlinkViewer is a user-friendly plugin for PyMOL v2 that maps intra-protein, inter-protein, and dead-end crosslinks onto protein structures/models and automates the calculation of inter-residue distances for the detected crosslinks. This enables rapid visualization of XL-MS data, assessment of whether a set of detected crosslinks is congruent with structural data, and easy production of high-quality images for publication.

KEYWORDS
chemical crosslinking, data visualization, mass spectrometry, PyMOL

1 | INTRODUCTION
Chemical crosslinking-mass spectrometry (XL-MS) is a powerful tool to derive structural information on proteins/protein assemblies, including transient or weak interactions necessary for biological function.1-3 XL-MS workflows begin by treatment of a purified protein/protein complex, lysate, or intact cells with an appropriate XL reagent. All crosslinkers have the same basic architecture, comprising two reactive groups (e.g., NHS-esters that primarily react with amines4), or radical-based crosslinkers (such as diazirines5 that react non-specifically) separated by a spacer arm.1-3 This spacer arm confers a distance constraint that can be used for comparison with high-resolution structures, structural modeling, model validation, or the detection of lowly populated states not captured by other methods.

After crosslinking, the proteins are digested (e.g., with trypsin) and the inter-protein, intra-protein, and dead-end (where only one of the two reactive groups react with...
the protein and the other remains unreacted or is quenched by solvent) crosslinked peptides must then be detected. These peptides are of relatively low abundance compared to their non-crosslinked counterparts, so much work has focused on developing analytical workflows to enable crosslink detection (e.g., using enrichment protocols, incorporating MS-cleavable groups in the spacer arm that yield diagnostic fragment ions, and developing specialized informatics tools). The detected crosslinks can be mapped onto protein sequences/networks (e.g., using xVis or xiNET) or protein structures for comparison with other data. Current tools for XL-MS data visualization include XLinkAnalyzer, a plugin for Chimera, as well as tools for the calculation and visualization of crosslinks as solvent accessible surface distance (SASD) paths (Xwalk, Jwalk, and DynamXL). However, no plugin exists for the commonly used PyMOL molecular graphics system.

Here, we present PyXlinkViewer, a tool for mapping chemical crosslinks onto a protein structure or structural model visualized within the PyMOL v2 molecular graphics system. PyXlinkViewer automates the display of inter-protein and intra-protein crosslinks and measures the Cα-Cα inter-residue Euclidean distances for each crosslink. These values can then be compared with the upper distance limit imposed by the crosslinker used. PyXlinkViewer also displays residues modified with dead-end crosslinks, which can probe for chemical accessibility and be used to score structural models; hence, our tool can also be used to easily visualize data from a range of other covalent labeling/footprinting methods (e.g., fast photochemical oxidation of proteins, carbene labeling, or other side-chain-specific probes). We envisage that this versatile tool will be a useful addition to XL-MS and other covalent labeling/MS workflows.

2 | RESULTS

We have developed the PyXlinkViewer plugin to enable rapid, simple visualization of inter-residue chemical crosslinks and dead-end crosslinks by mapping them onto a protein structure or structural model visualized within PyMOL v2. The source code, plugin installation file, example data, and user manual are freely available under a GNU General Public License at https://github.com/BobSchiffrin/PyXlinkViewer.

2.1 | Plugin design and installation

PyXlinkViewer is a cross-platform plugin for PyMOL v2 written in Python 3, which can be run on Linux, macOS, or Microsoft Windows operating systems. It has a dedicated graphical user-interface (GUI) (Figure 1) designed using QtDesigner v4.8.6. PyXlinkViewer is supplied as a ZIP file, with installation easily achieved using PyMOL’s plugin manager. A detailed user manual is available with the software. The code only imports modules present in the standard PyMOL v2 and Python 3 libraries, so no external dependencies need to be installed.

2.2 | Data import

After manual curation of an appropriate XL-MS dataset using dedicated XL-MS analysis software, the data are prepared as a list of crosslinks and dead-ends for use in PyXlinkViewer. This approach enables the user to curate their dataset and filter it based on the score of the detected crosslinks (each data analysis software package uses a different scoring algorithm) prior to data visualization. The file format required is that used by the existing tool Jwalk (see user manual) and allows maximum user control of the data displayed (see example data). This format details the residue numbers and chain identifiers of both residues involved in a crosslink. To visualize dead-end or covalent labeling/footprinting data, only one residue/chain identifier is included in each entry (see example data). There are a number of different file formats generated by the various XL-MS data processing tools. However, recently an XL-MS file conversion utility has been created, and a standard XL-MS data format proposed. The object-orientated design of PyXlinkViewer allows for simple extension to support a standard file format if adopted by the XL-MS community.

2.3 | Data visualization using PyXlinkViewer

After opening a data file a table of crosslinks is populated and the inter-residue Cα-Cα Euclidian distances are calculated and displayed (Figure 1). The upper distance threshold imposed by the spacer arm can be set, and the number of crosslinks that are satisfied/violated by this distance threshold is displayed. Crosslinks and dead-ends are drawn as lines between residue Cα atoms, or spheres at residue Cα atoms, respectively. Often residues are missing from the PDB files of protein structures (e.g., if they are disordered/unresolved by X-ray crystallography or cryogenic electron microscopy). If a residue involved in a crosslink is not present in the structure loaded into PyXlinkViewer, a warning message is printed in the PyMOL command window that the crosslink cannot be shown, which may draw attention to disordered regions involved in a protein interaction.
A number of visualization options can be changed in the GUI, including the color and width of the drawn crosslinks. The user can also show or hide different crosslink types (i.e., inter-protein, intra-protein, and dead-end). The XL table in the PyXlinkViewer GUI and the data visualized in the PyMOL display are automatically updated when the user makes any changes to the visualization options (see user manual). Each XL is drawn as a separate PyMOL object, labeled with the chain and residue IDs involved, so can be individually shown/hidden using the PyMOL GUI or command line. The data can be exported as a CSV file for further interrogation.

2.4 Visualizing XL-MS data for monomeric proteins and multimeric assemblies

To demonstrate the power of PyXlinkViewer, we highlight here its use with two publically available XL-MS datasets.

Firstly, we demonstrate the PyXlinkViewer workflow using intra-protein crosslinks detected in the chaperone SurA (Figure 2). SurA comprises three domains, a core domain (made of an N-terminal and C-terminal region), and two PPlase domains, P1 and P2 (Figure 2). SurA was crosslinked with the reagent disuccinimidyl dibutyric urea (DSBU), which contains two NHS-ester reactive groups (that primarily react with Lys residues and protein N-termini, but also with Ser, Thr, and Tyr), and has been shown to crosslink residues up to ~27–30 Å apart (Cα-Cα Euclidean distances) (21). Of the 32 intra-protein crosslinks detected, 13 were incompatible with the crystal structure when using a Euclidian distance cut-off of 27 Å, as determined using PyXlinkViewer. Eight crosslinks in SurA involved residues missing from the PDB structure (as detected by PyXlinkViewer), which were subsequently built using MODELER.22 As a PyMOL object is created for each crosslink, intra-domain and/or inter-domain crosslinks could easily be individually selected for display. While only 2/13 intra-domain crosslinks were incompatible with the crystal structure, most of the inter-domain crosslinks (11/19) were between residues greater than 27 Å apart. The data provided evidence that in solution, SurA populates conformations in which the P2 domain is much to the core and P1 domains than suggested by the crystal structure.23,24

In order to demonstrate PyXlinkViewer’s functions on a larger, multicomponent complex, we chose to visualize an XL-MS dataset obtained for the ORC-Cdc6-Cdt1-Mcm2-7 (OCCM) complex (Figure 3a), a helicase loading intermediate in DNA replication (14 subunits, ~1 MDa, 1,132 Lys-Lys crosslinks).25,26 On mapping the crosslinks to the cryoEM structure of OCCM,25 PyXlinkViewer identified that many of the crosslinks (625/1132) involved residues...
that were not present in the cryoEM structure. Of those that could be displayed, 77/507 were between residues greater than 27 Å apart in the structure. Using PyXlinkViewer’s ability to selectively display intra-chain or inter-chain XLs (Figure 3bc), it was easily possible to see that the violated intra-chain XLs are between the domains of multi-domain protein components of OCCM. Notably, violated intra-protein crosslinks were detected in the three proteins Cdc6, Mcm3, and Cdt1, suggesting some inter-domain motions/reorganization in these proteins in the complex [Figure 3c]. Additionally, it was easily discernible that many of the violated inter-protein XLs involved regions of different proteins within the complex that are not in defined secondary structure. Combined, these two examples of SurA and OCCM demonstrate that PyXlinkViewer can be used to visualize both large and small XL-MS datasets to rapidly gain structural/functional insights into proteins and their assemblies.

3 | DISCUSSION

Recent enhancements in XL-MS protocols,6,26,27 new crosslinker designs,28 and refined bioinformatics approaches8 have ensured that XL-MS is a key tool in the structural biologist’s armory of methods for determining the structure and dynamics of proteins and their complexes both in vitro and in vivo. Computational modeling methods using XL-MS restraints are being constantly improved,14–16,29 but a vital, final step in this pipeline is the comparison of structures/structural models with XL-MS data. This is necessary for model validation and data presentation. PyXlinkViewer is the first PyMOL plugin specifically designed to simply map crosslinks onto high-resolution structures/structural models, which is key for increasing the throughput of integrative structural studies. For reasons of speed and clarity, PyXlinkViewer calculates and displays Euclidean distances for each inter-residue XL. SASDs can be a more reliable indicator of whether or not a cross-link is satisfied,13,14 and routines, which take into account side-chain dynamics have also been developed.15 However, calculation of SASDs is currently approximately five orders of magnitude slower than for Euclidian distances,14 potentially prohibiting this approach when dealing with large numbers of crosslinks and/or structural models. The object-orientated design of PyXlinkViewer allows for extension of functionality in the future, including possible calculation and display of SASDs.

FIGURE 2  Visualization of SurA intra-protein crosslinks using PyXlinkViewer. SurA has a three-domain architecture, comprising the core domain (grey/orange), and two PPIase domains, P1 (green) and P2 (yellow). In the crystal structure of SurA (PDB: 1M5Y), the P2 domain is spatially separated from the core and P1 domains (23). (a) All detected crosslinks. (b) Intra-domain crosslinks. (c–e) Crosslinks detected between the (c) Core-P1, (d) Core-P2, and (e) P1-P2 domains. These data suggest that the P2 domain populates conformations in solution closer to the core and P1 domains than suggested by the crystal structure. Crosslinks are cultured blue or red, if they are satisfied or violated, respectively, using a Cα-Cα Euclidean distance threshold of 27 Å...
be displayed on the structure. This gives the user an opportunity to visually compare the distances of possible inter- and intra-subunit crosslinks for the same residue pair from different subunits and assess which are most compatible with the crosslinked peptides detected by MS. Experiments involving the detection of crosslinks between “light” (14N) and “heavy” (15N-labeled) subunits in a homooligomer can also assist in discriminating between inter- and intra-subunit crosslinks.30–32

PyXlinkViewer also enables visualization of dead-end crosslinks and data from other covalent labeling workflows (e.g., those that target specific amino acid residues,19 or radical labeling methods such as fast photochemical oxidation of proteins [FPOP] which labels promiscuously17,33). A number of software tools have been designed to enable the identification of residues that have been covalently labeled.34,35 The data from these analyses can easily be converted to a format compatible with import to PyXlinkViewer. However, it should be noted that for reagents/techniques that label many types of residues, for example, FPOP, it remains analytically challenging to determine at the residue level the site of the modification. Nevertheless, data from covalent labeling has shown promise as an input for structural modeling.26–30 To demonstrate the data input requirements for covalent labeling/dead-end XL data we have also included example data from FPOP analysis of the protein β2-microglobulin17 (see example data). While dead-end crosslinking data are often discarded from analysis, especially when XL-MS is used to aid structural modeling, there is emerging evidence of their importance as probes for chemical accessibility. Indeed, they can be used as restraints in modeling pipelines and as a tool for the scoring of structural models of proteins.16,36

In conclusion, PyXlinkViewer is a versatile plugin for PyMOL v2 that enables rapid interrogation of an XL-MS dataset by mapping crosslinks onto a structure and calculating their inter-residue distances. As PyXlinkViewer functions within PyMOL v2, publication quality figures depicting XL-MS data can be made with ease. We envisage that PyXlinkViewer will be a key tool for the expanding number of researchers using XL-MS to interrogate the functional mechanisms of proteins and protein assemblies.

ACKNOWLEDGMENTS
Members of the Radford and Brockwell Labs are thanked for helpful discussions. This work has been supported by the Biotechnology and Biological Sciences Research Council (BB/T000635/1, BB/P000037/1, and BB/N007603/1). A. N. C. acknowledges support from a University Academic Fellowship from the University of Leeds.

CONFLICT OF INTEREST
The authors declare no conflicts of interest.

AUTHOR CONTRIBUTIONS
Bob Schiffrin: Conceptualization; writing-original draft; writing-review and editing. Sheena Radford: Funding
acquisition; supervision; writing-review and editing.

**David Brockwell**: Funding acquisition; supervision; writing-review and editing.

**Antonio Calabrese**: Conceptualization; writing-original draft; writing-review and editing.

**DATA AVAILABILITY STATEMENT**

The source code, plugin installation file, test data, and user manual are freely available under a GNU General Public License at https://github.com/BobSchiffrin/PyXlinkViewer. PyMOL v2 is available from https://pymol.org/2/.

**ORCID**

Bob Schiffrin [https://orcid.org/0000-0002-1670-7356](https://orcid.org/0000-0002-1670-7356)

Sheena E. Radford [https://orcid.org/0000-0002-3079-8039](https://orcid.org/0000-0002-3079-8039)

David J. Brockwell [https://orcid.org/0000-0002-0802-5937](https://orcid.org/0000-0002-0802-5937)

Antonio N. Calabrese [https://orcid.org/0000-0003-2437-7761](https://orcid.org/0000-0003-2437-7761)

**REFERENCES**

1. Calabrese AN, Radford SE. Mass spectrometry-enabled structural biology of membrane proteins. Methods. 2018;147:187–205.
2. Chavez JD, Bruce JE. Chemical cross-linking with mass spectrometry: A tool for systems structural biology. Curr Opin Chem Biol. 2019;48:8–18.
3. Iacobucci C, Götzte M, Sinz A. Cross-linking/mass spectrometry to get a closer view on protein interaction networks. Curr Opin Biotech. 2020;63:48–53.
4. Kalkhof S, Sinz A. Chances and pitfalls of chemical cross-linking with amine-reactive N-hydroxysuccinimide esters. Anal Bioanal Chem. 2008;392:305–312.
5. Horne JE, Walko M, Calabrese AN, et al. Rapid mapping of protein interactions using tag-transfer photocrosslinkers. Angew Chem Int Ed. 2018;57:16688–16692.
6. Schmidt R, Sinz A. Improved single-step enrichment methods of cross-linked products for protein structure analysis and protein interaction mapping. Anal Bioanal Chem. 2017;409:2393–2400.
7. Sinz A. Divide and conquer: Cleavable cross-linkers to study protein conformation and protein–protein interactions. Anal Bioanal Chem. 2017;409:33–44.
8. Yilmaz S, Shiferaw GA, Rayo J, Economou A, Martens L, Vandermaarliere E. Cross-linked peptide identification: A computational forest of algorithms. Mass Spectrom Rev. 2018;37:738–749.
9. 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:W362–W369.
10. Combe CW, Fischer L, Rappsilber J. xiNET: Cross-link network maps with residue resolution. Mol Cell Proteomics. 2015;14:1137–1147.
11. Kosinski J, von Appen A, Ori A, Karius K, Muller CW, Beck M. Xlink analyzer: Software for analysis and visualization of cross-linking data in the context of three-dimensional structures. J Struct Biol. 2015;189:177–183.
12. Pettersen EF, Goddard TD, Huang CC, et al. UCSF chimera—a visualization system for exploratory research and analysis. J Comput Chem. 2004;25:1605–1612.
13. Kahraman A, Malmström L, Aebersold R. Xwalk: Computing and visualizing distances in cross-linking experiments. Bioinformatics. 2011;27:2163–2164.
14. Bullock JMA, Thalassinos K, Topf M. Jwalk and MNXL web server: Model validation using restraints from crosslinking mass spectrometry. Bioinformatics. 2018;34:3584–3585.
15. Degiacomi MT, Schmidt C, Baldwin AJ, Beneschi JLP. Accommodating protein dynamics in the modeling of chemical cross-links. Structure. 2017;25:1751–1757.
16. Sinnott M, Malhotra S, Madhusudhan MS, Thalassinos K, Topf M. Combining information from crosslinks and monolinks in the modelling of protein structures. Structure. 2020;28:1–10.
17. Cornwell O, Radford SE, Ashcroft AE, Ault JR. Comparing hydrogen deuterium exchange and fast photochemical oxidation of proteins: A structural characterisation of wild-type and deltaN6 beta2-microglobulin. J Am Soc Mass Spectrom. 2018;29:2413–2426.
18. Manzi L, Barrow AS, Scott D, et al. Carbene footprinting accurately maps binding sites in protein-ligand and protein-protein interactions. Nat Commun. 2016;7:13288.
19. Mendoza VL, Vacht RW. Probing protein structure by amino acid-specific covalent labeling and mass spectrometry. Mass Spectrom Rev. 2009;28:785–815.
20. Bender J, Schmidt C. The CroCo cross-link converter: A user-centred tool to convert results from cross-linking mass spectrometry experiments. Bioinformatics. 2019;36:1296–1297.
21. Merkley ED, Rysavy S, Kahraman A, Hafen RP, Daggett V, Adkins JN. Distance restraints from crosslinking mass spectrometry: Mining a molecular dynamics simulation database to evaluate lysine-lysine distances. Protein Sci. 2014;23:747–759.
22. Eswar N, Webb B, Marti-Renom MA, et al. Comparative protein structure modeling using modeller. Current Protoc Bioinf. 2006, 15:5.6.1–5.6.30. http://dx.doi.org/10.1002/0471250953.bi0506s15.
23. Bitto E, McKay DB. Crystallographic structure of SurA, a molecular chaperone that facilitates folding of outer membrane porins. Structure. 2002;10:1489–1498.
24. Calabrese AN, Schiffrin B, Watson M, et al. Inter-domain dynamics in the chaperone SurA and multi-site binding to its unfolded outer membrane protein clients. Nat Comm. 2020;11:2155.
25. Yuan Z, Riera A, Bai L, et al. Structural basis of Mcm2–7 replicative helicase loading by ORC–Cdc6 and Cdt1. Nat Struct Mol Biol. 2017;24:316–324.
26. Mendes ML, Fischer L, Chen ZA, et al. An integrated workflow for crosslinking mass spectrometry. Mol Syst Biol. 2019;15:8994.
27. Iacobucci C, Götzte M, Ihling CH, et al. A cross-linking/mass spectrometry workflow based on MS-cleavable cross-linkers and the MeroX software for studying protein structures and protein–protein interactions. Nat Protocols. 2018;13:2864–2889.
28. Steigenberger B, Pieters RJ, Heck AJR, Scheltema RA. PhoX: An IMAC-Enrichable cross-linking reagent. ACS Central Sci. 2019;5:1514–1522.
29. Ding YH, Gong Z, Dong X, et al. Modeling protein excited-state structures from “over-length” chemical cross-links. J Biol Chem. 2017;292:1187–1196.
30. Taverner T, Hall NE, O’Hair RAJ, Simpson RJ. Characterization of an antagonist interleukin-6 dimer by stable isotope labeling, cross-linking, and mass spectrometry. J Biol Chem. 2002;277:46487–46492.
31. Pettelkau J, Thondorf I, Theisgen S, et al. Structural analysis of guanylyl cyclase-activating protein-2 (GCAP-2) homodimer by stable isotope-labeling, chemical cross-linking, and mass spectrometry. J Am Soc Mass Spectrom. 2013;24:1969–1979.
32. Arlt C, Ihling CH, Sinz A. Structure of full-length p53 tumor suppressor probed by chemical cross-linking and mass spectrometry. Proteomics. 2015;15:2746–2755.
33. Calabrese AN, Ault JR, Radford SE, Ashcroft AE. Using hydroxyl radical footprinting to explore the free energy landscape of protein folding. Methods. 2015;89:38–44.
34. Bellamy-Carter J, Oldham NJ. PepFoot: A software package for semiautomated processing of protein footprinting data. J Proteome Res. 2019;18:2925–2930.
35. Ziemianowicz DS, Sarpe V, Schriemer DC. Quantitative analysis of protein covalent labeling mass spectrometry data in the mass spec studio. Anal Chem. 2019;91:8492–8499.
36. Schmidt C, Macpherson JA, Lau AM, Tan KW, Fraternali F, Politis A. Surface accessibility and dynamics of macromolecular assemblies probed by covalent labeling mass spectrometry and integrative modeling. Anal Chem. 2017;89:1459–1468.
37. Limpikirati P, Liu T, Vachet RW. Covalent labeling-mass spectrometry with non-specific reagents for studying protein structure and interactions. Methods. 2018;144:79–93.
38. Aprahamian ML, Lindert S. Utility of covalent labeling mass spectrometry data in protein structure prediction with Rosetta. J Chem Theory Comput. 2019;15:3410–3424.

How to cite this article: Schiffrin B, Radford SE, Brockwell DJ, Calabrese AN. PyXlinkViewer: A flexible tool for visualization of protein chemical crosslinking data within the PyMOL molecular graphics system. Protein Science. 2020;29:1851–1857. https://doi.org/10.1002/pro.3902