

xiView: A common platform for the downstream analysis of Crosslinking Mass Spectrometry data

Martin Graham^{a,*}, Colin Combe^{a,*}, Lars Kolbowski^{a,b} and Juri Rappsilber^{a,b}

a. Wellcome Centre for Cell Biology, University of Edinburgh, Edinburgh EH9 3BF, United Kingdom.

b. Bioanalytics, Institute of Biotechnology, Technische Universität Berlin, 13355 Berlin, Germany.

* equal contribution

Abstract

xiView provides a common platform for the downstream analysis and visualisation of Crosslinking Mass Spectrometry data. It is independent of the search software used and its input is compliant with the relevant mass spectrometry data standards. It uses established visualisation techniques, notably Multiple Coordinated Views, to help the user explore the data and is designed to facilitate comparisons between different datasets.

Main

Crosslinking Mass Spectrometry (CLMS) investigates the structure and function of proteins by providing proximity information for pairs of amino acids¹. xiView is a web-based tool for exploring these datasets, as are XLink-DB², xVis³, ProXL⁴ and CLMSVault⁵. However, xiView is distinguished from these by (i) standards compliant input data, and (ii) simultaneous interactive multiple views that are linked together, each giving a different yet synchronised perspective on the dataset.

The supported data standards are HUPO-PSI mzIdentML⁶ and HUPO-PSI mzML⁷, with two non-standard formats also supported - a peak list format (.mgf⁸) and one for identifications (CSV). The software has been proven to work with valid mzIdentML, using output from SIM-XL¹⁷ as an example.

The typical graphical representations of data to be found in both CLMS publications and CLMS software are: a) annotated spectra; b) 2D network representations such as XiNet⁹ and Circos¹⁰-style views; c) 3D structures with crosslinks mapped onto them; and d) histograms of link distances derived from the 3D structure. xiView expands on this collection with e) a matrix / contact map for chosen protein pairs; f) a protein information panel showing a protein sequence as a wrapped string with crosslinks marked within it; g) visual sequence alignments between search sequences and the chosen PDB file which are partially adjustable; h) a scatterplot; and i) a general-purpose non-distance histogram. Both the scatterplot and histogram display metadata about peptide spectrum matches (PSMs) and/or crosslinks (unique distance restraints) - the specific attributes are selected by the user. All the views along with supporting legends can be exported to figure-quality SVG (Scalable Vector Graphics) files or PNG in the case of the 3D view.

Unlike similar tools, xiView combines all these representations in a single web page and they are all accessible simultaneously. Colour schemes can be chosen based on crosslink attributes and are applied across all views. Crosslinks are filtered using conditions set in a filter bar with the resulting set of crosslinks immediately reflected across all views. Crosslinks selected or highlighted in one view are automatically selected or highlighted in all other open views. This design is in-keeping with the established data visualisation technique known as ‘Multiple Coordinated Views’¹¹ - each individual view accentuates and reveals a different aspect of the data, and the coordination between them aids in understanding the relationships between these aspects, which ultimately leads to a better understanding of the dataset.

Across the crosslinking community, it is accepted that *“the term ‘cross-link’ refers to the specific amino acid residues that are connected, irrespective of different peptide sequences due to missed cleavage sites or modifications.”*¹² A PSM provides a single distance restraint by identifying a pair of crosslinked peptides. Typically, this has been in one spectrum, though isotope labelling and multiple fragmentation techniques can mean a PSM is associated with more than one spectrum (see Section 5.2.9 of the MzIdentML 1.2 specification). There may be several different interpretations of the same spectrum, though search software generally only allows one interpretation of the same spectrum to pass its validation threshold.

xiView uses xiSPEC¹³ for the display of annotated spectra, including the option to view alternative explanations of the same spectrum. Like XLink-DB² (v2) and CLMSVault⁵, xiView uses xiNET⁹ to display a 2D network of the crosslinks. xiNET and the other network views in xiView can also represent ambiguities that result when a peptide occurs in more than one location in the search sequence database.

Uniprot annotations are automatically acquired when search proteins are assigned a valid UniProtKB accession number. A number of types of metadata can also be loaded into xiView. When a PDB¹⁴ structure is loaded the crosslinks are aligned onto the 3D structures and shown using NGL¹⁵. The matrix view also shows distances derived from the structure as a shaded background contact map. Often the search sequences do not match the canonical Uniprot sequences, and PDB sequences often offer only partial coverage of sequences. Integrating the Uniprot and PDB data thus requires the calculation of sequence alignments, performed internally within xiView, which are then used to ensure annotations in the xiNet and circular views, and crosslinks and distance measurements in the NGL component, are mapped to the correct place on proteins.

A typical session with xiView is then as follows: an individual or set of crosslink datasets are chosen by a user to view - xiView allows multiple datasets to be aggregated and/or compared through assigning group identifiers. Filtered crosslinks are shown in the collection of co-ordinated views. When crosslinks within one of these views are selected, the corresponding representations of those crosslinks are then accented in the other views to reveal possible patterns e.g. an area of the scatterplot view can be selected to pick crosslinks with a certain number of PSMs and scores for those PSMs, and the selection seen in the circular or network view to see if there is a positional pattern. At the same time, the supporting PSMs and data are shown in a table, and selecting an individual PSM within that table then shows the corresponding annotated spectrum in another view, see Figure 1. Then, using

the metadata loading option, an appropriate PDB file can be loaded, and displayed to assess filtered and selected crosslinks in the context of the physical protein structure. Finally, filtered sets of cross-links and individual views can be exported.

xiView's capabilities have evolved according to the demands of a set of real scientific users and have also undergone testing to discover and correct usability issues, and thus can be confidently said to support real user tasks in inspecting crosslinking data sets.

Documentation for the software can be found at <https://xiview.org/xidocs/html/xiview.html>, and short tutorial videos can be found at <http://rappsilberlab.org/rappsilber-laboratory-home-page/tools/xiview/xiview-videos/>.

A public facing version of the software is available at <https://xiview.org> where example data sets can be explored and new datasets can be uploaded and explored. There is a file size limit of 1GB for uploads. For datasets with a file sizes larger than 1GB, or where users do not wish their data stored by a third party, instructions for a local installation of xiView are available at https://github.com/xiView_container.

Code availability

Source code is available at https://github.com/xiView_container. This code is published under the open source license GPL v3.

Acknowledgments

We acknowledge the work of Alex Rose whose NGL 3D viewer¹⁵ is embedded as the basis of one of the views within xiView and Lutz Fischer, the author of the xiAnnotator which xiSPEC¹³ uses to annotate the spectra. Thanks to Zhuo Chen for the use of the example data set. This work was supported by the Wellcome Trust through a Senior Research Fellowship to J.R. [103139]. The Wellcome Centre for Cell Biology is supported by core funding from the Wellcome Trust [203149].

Author Contributions

M.G., C.C., and L.K. wrote the software, J.R. conceived of the project. All authors contributed to writing the manuscript.

References

1. Rappsilber, J. The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. *J. Struct. Biol.* **173**, 530–540 (2011).

2. Zheng, C. *et al.* XLink-DB: database and software tools for storing and visualizing protein interaction topology data. *J. Proteome Res.* **12**, 1989–1995 (2013).
3. 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.* **43**, W362–9 (2015).
4. Riffle, M., Jaschob, D., Zelter, A. & Davis, T. N. ProXL (Protein Cross-Linking Database): A Platform for Analysis, Visualization, and Sharing of Protein Cross-Linking Mass Spectrometry Data. *J. Proteome Res.* **15**, 2863–2870 (2016).
5. Courcelles, M. *et al.* CLMSVault: A Software Suite for Protein Cross-Linking Mass-Spectrometry Data Analysis and Visualization. *J. Proteome Res.* **16**, 2645–2652 (2017).
6. Vizcaíno, J. A. *et al.* The mzIdentML Data Standard Version 1.2, Supporting Advances in Proteome Informatics. *Mol. Cell. Proteomics* **16**, 1275–1285 (2017).
7. Martens, L. *et al.* mzML--a community standard for mass spectrometry data. *Mol. Cell. Proteomics* **10**, R110.000133 (2011).
8. Deutsch, E. W. File formats commonly used in mass spectrometry proteomics. *Mol. Cell. Proteomics* **11**, 1612–1621 (2012).
9. Combe, C. W., Fischer, L. & Rappsilber, J. xiNET: cross-link network maps with residue resolution. *Mol. Cell. Proteomics* **14**, 1137–1147 (2015).
10. Krzywinski, M. *et al.* Circos: An information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).
11. Roberts, J. C. State of the Art: Coordinated amp; Multiple Views in Exploratory Visualization. in *Fifth International Conference on Coordinated and Multiple Views in Exploratory Visualization (CMV 2007)* 61–71 (2007).
12. Iacobucci, C. *et al.* Cross-linking/Mass Spectrometry: A Community-Wide, Comparative Study Towards Establishing Best Practice Guidelines. *bioRxiv* 424697 (2018). doi:10.1101/424697
13. Kolbowski, L., Combe, C. & Rappsilber, J. xiSPEC: web-based visualization, analysis and sharing of proteomics data. *Nucleic Acids Res.* **46**, W473–W478 (2018).
14. Berman, H., Henrick, K. & Nakamura, H. Announcing the worldwide Protein Data Bank. *Nat. Struct. Biol.* **10**, 980 (2003).
15. Rose, A. S. & Hildebrand, P. W. NGL Viewer: a web application for molecular visualization. *Nucleic Acids Res.* **43**, W576–9 (2015).
16. UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Res.* **43**, D204–12 (2015).
17. Lima, D. B. *et al.* SIM-XL: A powerful and user-friendly tool for peptide cross-linking analysis. *J. Proteomics* **129**, 51–55 (2015).

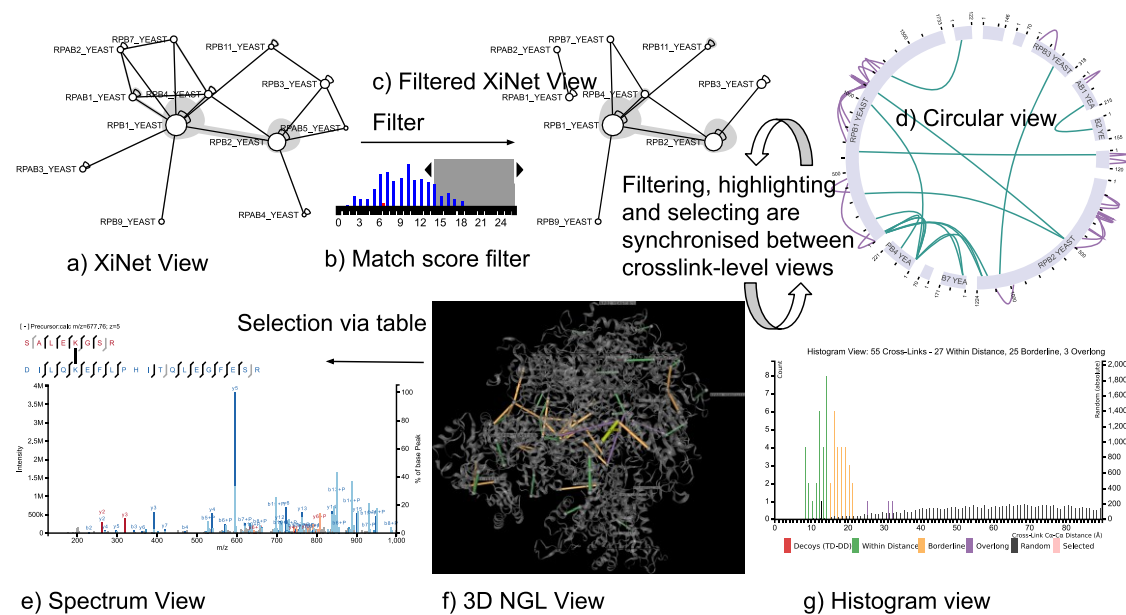


Figure 1. Some of the interactive components of XiView. Clockwise from top left: a) Initial network view, b) visual filter for match scores, c) the resulting filtered network view, d) circular view with same filtering. Upon loading of a PDB file we can see g) a histogram view showing distance against random background with a distance colour scheme, f) a 3D NGL view with crosslinks automatically aligned to PDB sequences and a distance colour scheme. Finally we can interrogate specific PSM details in e) the spectrum view. Selected links populate a table of PSM's and spectra is launched from there. Views a) d) f) and g) all respond immediately to changes in the filter, along with user selections and highlighting.