

Bioinformatics, YYYY, 0–0

doi: 10.1093/bioinformatics/xxxxx

Advance Access Publication Date: DD Month YYYY

Manuscript Category

Application Note

Jwalk and MNXL Web Server: Model Validation using Restraints from Crosslinking Mass Spectrometry

JMA Bullock¹, K Thalassinos^{1,2} and M Topf^{1*}¹Institute for Structural and Molecular Biology, Birkbeck College, University of London, ²Institute for Structural and Molecular Biology, UCL, University of London

*To whom correspondence should be addressed.

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Motivation: Crosslinking Mass Spectrometry generates restraints that can be used to model proteins and protein complexes. Previously, we have developed two methods, to help users achieve better modelling performance from their crosslinking restraints: Jwalk, to estimate solvent accessible distances between crosslinked residues and MNXL, to assess the quality of the models based on these distances.

Results: Here we present the Jwalk and MNXL webservers, which streamline the process of validating monomeric protein models using restraints from crosslinks. We demonstrate this by using the MNXL server to filter models made of varying quality, selecting the most native-like.

Availability: The webserver and source code are freely available from jwalk.ismb.lon.ac.uk and mnxl.ismb.lon.ac.uk.

Contact: m.topf@cryst.bbk.ac.uk, j.bullock@cryst.bbk.ac.uk

1 Introduction

Crosslinking Mass Spectrometry (XL-MS) is an experimental method that can generate sparse structural information on proteins, complementary to traditional structural techniques. Briefly, an XL-MS experiment consists of (a) crosslinking of your target protein (or protein complex); (b) digesting the crosslinked protein; and (c) identifying via MS which residues are crosslinked. This results in restraining information that can then be used to model the protein of interest.

Jwalk (Bullock et al., 2016) is a program that calculates the Solvent Accessible Surface Distance (SASD), which is defined as the shortest distance between two residues across the surface of the protein (Kahraman et al., 2011). The SASD is a theoretically more correct approximation of the distance between crosslinked residues than the commonly used Euclidean distance, because the Euclidean distance permits travelling through the protein mass whereas a crosslinker cannot travel through protein mass. Other methods exist that calculate the SASD, while also considering side-chain flexibility (Degiacomi et al., 2017).

Crosslinks can also act as a proxy for solvent accessibility, as residues must be solvent exposed if they are to be crosslinked. Utilising

this extra information, we created the crosslink scoring function called *Matched and Non-accessible Crosslink Score* (MNXL), which we found to outperform more conventional methods when scoring protein monomers (Bullock et al., 2016). In order to streamline the XL-MS modelling procedure, we have incorporated these two developments (MNXL and Jwalk) into two webservers. Both programs are also standalone and freely available to download.

2 Implementation

Jwalk and MNXL are written in Python 2.7. For a full description of Jwalk see (Bullock et al., 2016). Jwalk outputs a list of all the SASDs and Euclidean distances between target residues in a *.txt* file along with a *.pdb* file that contains all the SASD paths modelled using glycine pseudo atoms. On the Jwalk webserver (Figure 1a), SASD paths are visualised with JSmol (Hanson et al., 2013). MNXL takes as input a list of experimental crosslinks and Jwalk *.txt* output files (which can also be provided by the user independently). The SASDs of the experimental crosslinks are then scored using the MNXL scoring function (Bullock et al., 2016). Using the webserver, users can also go directly from *.pdb* file to MNXL score (instead of running Jwalk separately).

© The Author(s) 2018. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License

(<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the

original work is properly cited.

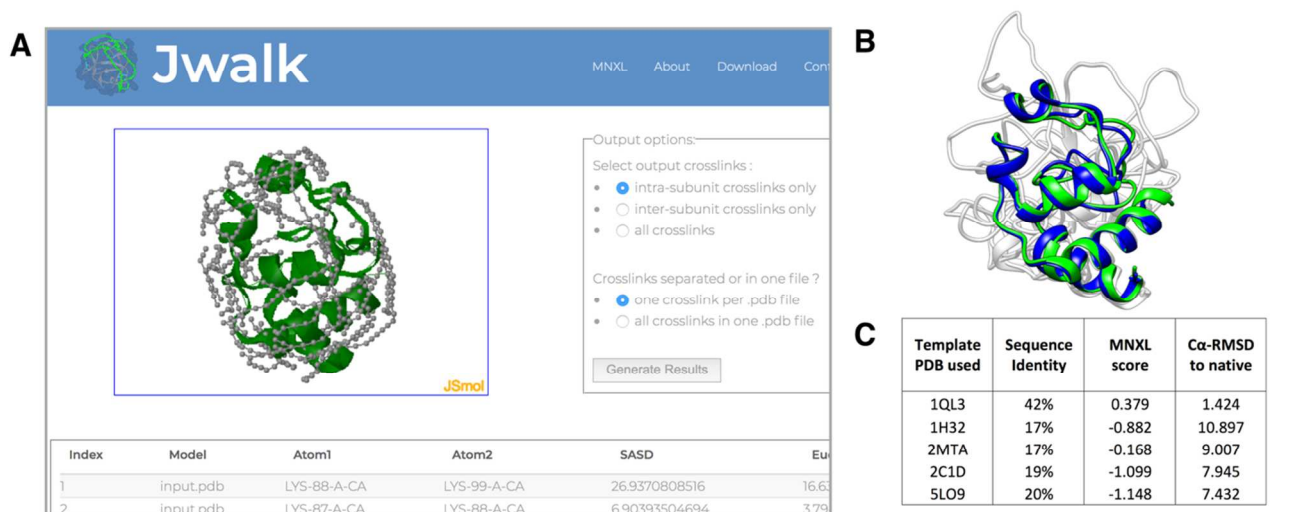


Figure 1 [(A). Partial screenshot of the Jwalk result page with PDB id. 1HRC shown. (B) The models of the test-case superposed in grey, with the best scoring model based on 1QL3 (blue) and the native 1HRC (green). (C) Results table showing MNXL is able to select the lowest C α -RMSD

MNXL outputs the scores for each model in a *.txt* file. Higher scores indicate better models. Additionally, MNXL outputs the number of crosslinks that are matched, violating and non-accessible to aid model assessment. The source code for both Jwalk and MNXL is available under a Creative Commons license at <http://topf-group.ismb.lon.ac.uk/Software.html>.

3 Results

To demonstrate the utility of the Jwalk/MNXL web server, we show the ability for the combination of Jwalk and MNXL to filter homology models made with different templates. 5 different comparative models of the horse heart cytochrome C crystal structure (PDB id: 1HRC) were made with MODELLER (Eswar et al., 2007) using templates of various quality, taken from the HHPred server (Alva et al., 2016) with probability score > 96% in all cases (Figure 1b). The five template PDB ids (and associated sequence identity) are: 1QL3 (42%), 5LO9 (20%), 1H32 (17%), 2MTA (17%) and 2C1D (19%), respectively. These PDB models were then uploaded into the MNXL webserver (along with the experimentally observed crosslinks taken from XLdb (Kahraman et al., 2013)). The MNXL score was able to successfully select the model made using template 1QL3, which is the model with the lowest C α -RMSD to the native structure (Figure 1c) – for further discussion of the results see Supplementary Information.

4 Conclusion

We have created webservers for MNXL and Jwalk, two methods that can be used to validate models using restraints from crosslinking mass spectrometry and demonstrated how it can be useful in filtering homology models built from different templates. These webservers are designed to be user-friendly, in order to make it easier for the novice user to make better use of their crosslinking data. We hope to expand these platforms to incorporate the modelling of protein complexes in the near future.

Acknowledgements

We thank Dr. David Houldershaw for computer support and Dr. Agnel Praveen-Joseph, the Topf and Thalassinos groups for helpful discussions.

Funding

This work was supported by BBSRC London Interdisciplinary Doctoral Programme (to J.B) and MRC MR/M019292/1 (to M.T.)

References

- Alva, V., Nam, S.-Z., Söding, J., and Lupas, A.N. (2016). The MPI bioinformatics Toolkit as an integrative platform for advanced protein sequence and structure analysis. *Nucleic Acids Res.* *44*, gkw348.
- Bullock, J.M.A., Schwab, J., Thalassinos, K., and Topf, M. (2016). The importance of non-accessible crosslinks and solvent accessible surface distance in modelling proteins with restraints from crosslinking mass spectrometry. *Mol. Cell. Proteomics* *15*, 2491–2500.
- Degiacomis, M.T., Schmidt, C., Baldwin, A.J., and Benesch, J.L.P. (2017). Accommodating Protein Dynamics in the Modeling of Chemical Crosslinks. *Structure* *25*, 1751–1757.e5.
- Eswar, N., Webb, B., Marti-Renom, M. a, Madhusudhan, M.S., Eramian, D., Shen, M.-Y., Pieper, U., and Sali, A. (2007). Comparative protein structure modeling using MODELLER. *Current protocols in protein science* (Unit 2.9)
- Hanson, R.M., Prilusky, J., Renjian, Z., Nakane, T., and Sussman, J.L. (2013). JSmol and the next-generation web-based representation of 3D molecular structure as applied to proteopedia. *Isr. J. Chem.* *53*, 207–216.
- Herzog, F., Kahraman, A., Boehringer, D., Mak, R., Bracher, A., Walzthoeni, T., Leitner, A., Beck, M., Hartl, F.-U., Ban, N., et al. (2012). Structural Probing of a Protein phosphatase 2A network by chemical cross-linking and mass spectrometry. *Science* (80-.). *9114*, 1348–1352.
- Kahraman, A., Malmström, L., and Aebersold, R. (2011). Xwalk: computing and visualizing distances in cross-linking experiments. *Bioinformatics* *27*, 2163–2164.
- Kahraman, A., Herzog, F., Leitner, A., Rosenberger, G., Aebersold, R., and Malmström, L. (2013). Cross-link guided molecular modeling with ROSETTA. *PLoS One* *8*, e73411.
- Mitternacht, S. (2016). FreeSASA: An open source C library for solvent accessible surface area calculations. *F1000Research* *5*, 189.
- Mizuguchi, K., Deane, C.M., Blundell, T.L., Johnson, M.S., and Overington, J.P. (1998). JOY: Protein sequence-structure representation and analysis. *Bioinformatics* *14*, 617–623.