Bioinformatics, 37(18), 2021, 3035–3037
doi: 10.1093/bioinformatics/btab149
Advance Access Publication Date: 3 March 2021
Applications Note

Structural bioinformatics

Characterizing protein conformers by cross-linking mass spectrometry and pattern recognition

Louise U. Kurt[1,*,†], Milan A. Clasen[1,†], Marlon D. M. Santos[1], Eduardo S. B. Lyra[2], Luana O. Santos[2], Carlos H. I. Ramos[2], Diogo B. Lima[3], Fabio C. Gozzo[2] and Paulo C. Carvalho[1,*]

[1]Laboratory for Structural and Computational Proteomics, Carlos Chagas Institute, Fiocruz, Paraná 81350-010, Brazil[2]Institute of Chemistry, University of Campinas, São Paulo 13083-862, Brazil and [3]Department of Chemical Biology, Leibniz – Forschungsinstitut für Molekulare Pharmakologie (FMP), Berlin 13125, Germany

*To whom correspondence should be addressed.
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.
Associate Editor: Pier Luigi Martelli

Received on November 19, 2020; revised on February 25, 2021; editorial decision on February 27, 2021; accepted on March 2, 2021

Abstract

Motivation: Chemical cross-linking coupled to mass spectrometry (XLMS) emerged as a powerful technique for studying protein structures and large-scale protein-protein interactions. Nonetheless, XLMS lacks software tailored toward dealing with multiple conformers; this scenario can lead to high-quality identifications that are mutually exclusive. This limitation hampers the applicability of XLMS in structural experiments of dynamic protein systems, where less abundant conformers of the target protein are expected in the sample.
Results: We present QUIN-XL, a software that uses unsupervised clustering to group cross-link identifications by their quantitative profile across multiple samples. QUIN-XL highlights regions of the protein or system presenting changes in its conformation when comparing different biological conditions. We demonstrate our software's usefulness by revisiting the HSP90 protein, comparing three of its different conformers. QUIN-XL's clusters correlate directly to known protein 3D structures of the conformers and therefore validates our software.
Availabilityand implementation: QUIN-XL and a user tutorial are freely available at http://patternlabforproteomics.org/quinxl for academic users.
Contact: lulrichkurt@gmail.com
Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Chemical cross-linking coupled to mass spectrometry (XLMS) has become a powerful approach for the structural characterization of protein systems. The XLMS approach usually consists of covalently stabilizing a protein or protein complex with a cross-linker molecule, followed by enzymatic digestion and analysis via tandem mass spectrometry. The results ultimately provide spatial constraints revealing key structural information, such as protein folding, the topology of complexes and the interaction region between proteins, among others (Sinz *et al.*, 2015; Chavez and Bruce, 2019). Recent works have gone as far as attempting full *ab-initio* modeling from quality cross-linked peptide-pair identifications (XL-IDs) alone, with impressive results (Mintseris and Gygi, 2020; Kahraman *et al.*, 2013).

Acquiring XL-IDs that can aid in structural information, however, is not always straightforward. The relatively low intensity of acquired XL spectra and the increased difficulties in correctly identifying cross-links, as compared to linear peptides, makes selecting confident search results more challenging, often requiring novel algorithms (Iglesias *et al.*, 2010; Fischer and Rappsilber, 2017). In particular, for experiments aiming to structurally characterize proteins or complexes, target proteins in the sample may be present in different conformers, and isolating a single target can be complicated (Phillips *et al.*, 2015). This may lead to XL-IDs that cannot, at first, be assigned to a specific conformation. As far as we know, there is no established approach to deal with the presence of multiple conformers.

We present QUIN-XL, a software that deals with the multiple conformer problem by leveraging quantitative variations of XL-IDs across multiple treatments for clustering identifications by quantitative profile. Traditional shotgun proteomics has presented similar approaches (de Saldanha da Gama Fischer *et al.*, 2010), but none directly applied to XLMS data analysis. QUIN-XL provides robust

XIC quantitation and visualization, along with an XL-ID clustering module, which functions by encoding quantitative values of identified species into vectors, henceforth referred to as quantitative profiles, that are clustered according to similar abundance patterns. Selected XL-IDs can be exported to other structural XLMS software, such as PyMOL, developed by DeLano Scientific (DeLano, 2002) and Topolink (Ferrari *et al.*, 2019). Finally, all this is built within a friendly graphical user interface. We demonstrate the effectiveness of our approach by characterizing three conformers of the 90 kDa Heat Shock Protein (HSP90).

2 Materials and methods

Human HSP90 was enriched for three biological conditions: ATP-bound, ADP-bound and in the Apo form; we note that the enrichment still allowed for other conformers to be found in the same sample. The proteins for each biological condition were independently subjected to the cross-linking reaction with DSS, trypsinized and analyzed using an Orbitrap Fusion Lumos Mass Spectrometer (ThermoFisher Scientific). XL-ID search was performed by the SIM-XL software (Lima *et al.*, 2018).

QUIN-XL reads cross-link identification files directly from the SIM-XL search engine, or from the community standard mzIdentML 1.2 (Vizcaíno *et al.*, 2017) file formats. Subsequently, condition IDs are assigned to each biological condition; technical replicates are treated as the same biological state in the software. Then, QUIN-XL acquires extracted ion chromatograms (XIC) for all XL-IDs. Manual user curation and several post-quantification filters are available, both at the identification and the XIC curve level.

QUIN-XL performs XL-ID clustering according to their quantitative profiles. Each cross-link is encoded as a vector of n dimensions (where n is the number of condition IDs initially assigned), and with the respective quantitation values as arguments for the vectors. These vectors serve as input for the k-means algorithm (Lloyd, 1982). Optimal number of clusters is automatically determined by performing multiple clustering runs while varying k, utilizing the Silhouette method to evaluate cluster groups (Rousseeuw, 1987). The resulting clusters are shown on a user-friendly graphical interface, with a 2D-Map representation of the protein with the XLs from each cluster available to facilitate identifying regions that differ the most between possible conformations. The results can also be exported for visualization on PyMOL and for further analysis on Topolink.

3 Results

One hundred and seventeen confident XL-IDs were reported by SIM-XL. QUIN-XL's clustering approach determined an optimal number of four clusters, where three of them contained XLs mostly exclusive to one of the three biological conditions, with the fourth cluster containing XLs common to all conditions (Fig. 1A) (further discussion in Supplementary File—QUIN-XL Clustering). Existing 3D models of the HSP90 conformers were used to verify the robustness of our results. QUIN-XL correctly characterized all three HSP90 conformers and highlighted the regions with structural alterations.

Analysis of the clusters suggested that the middle and C-terminal domains of the HSP90 present significant differences between the conformers, which is expected, as these regions refer to the anchoring site of the protein's open/close movement. Our results also listed changes in the N-terminal domain in the Apo and ADP-bound states, as the closed conformation of the ATP-bound conformer is less exposed to the solvent (Fig. 1B and C). Visual inspection of the XL-ID clusters also indicates the existence of other conformers in the samples. For instance, some XLs were found in all conditions, yet they were considerably more abundant in the Apo state. This leads to the conclusion that, due to the dynamic equilibrium of the protein, the Apo conformer is also significantly present in other conditions. The use of QUIN-XL makes these conclusions almost immediate, providing valuable insight into protein and protein complexes

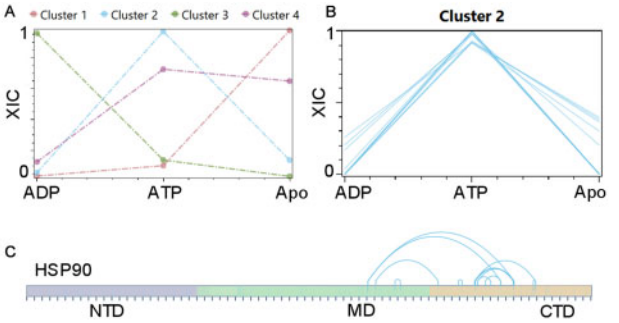

Fig. 1. QUIN-XL's results. (**A**) Proteomic quantitative profiles for the clusters identified by QUIN-XL, representing four different quantitative profiles in the HSP90 experiment. *y*-axis is the normalized XIC values of the XL-IDs along with the different biological conditions in the *x*-axis; (**B**) Graphical representation of one of the clusters generated, where each blue line represents the quantification profile of a single XL-ID. Axes are the same as in (A). In this case, the cluster shows XLs that were mostly exclusive to the ATP-bound sample, probably belonging to the most abundant conformer in that sample; (**C**) 2D-Map of HSP90 [in purple is the N-terminal domain (NTD), in green is the middle domain (MD) and in orange is the C-terminal domain (CTD)] with the cross-links from the cluster in (B), showing a region of very pointed differences between this conformer and the ones in other samples. (Color version of this figure is available at *Bioinformatics* online.)

while also supplying potentially valuable separation of XL-IDs for downstream modeling.

We further demonstrate the effectiveness of QUIN-XL on a public dataset by Jung *et al.*, (2020) (ProteomeXchange ID: PXD013907). The authors studied the structural modifications to the huntingtin (HTT) protein caused by changes in the polyglutamine region and by alterations in the phosphorylation pattern. Cross-linking experiments assessed the structure of HTT in four biological conditions [HTT with 23 glutamines (Q23-HTT), HTT with 78 glutamines (Q78-HTT), Q23-HTT with a mutation on Ser2116 (Q23-S2116A) and Q78-HTT with a mutation on Ser2116 (Q78-S2116A)]. QUIN-XL's corroborated the original paper's findings, particularly pinpointing several XL-IDs enriched for the Q78-HTT conformer. Additional documentation on our findings for this matter is available in the Supplementary File.

Funding

This work was supported by Fiocruz, the Brazilian National Research Council (CNPq—Universal) and the Graduate Studies Agency (CAPES).

Conflict of Interest: none declared.

Data availability

Data are available via ProteomeXchange with identifier PXD022443.

References

Chavez,J.D. and Bruce,J.E. (2019) Chemical cross-linking with mass spectrometry: a tool for systems structural biology. *Curr. Opin. Chem. Biol.*, **48**, 8–18.

de Saldanha da Gama Fischer J. *et al.* (2010) Dynamic proteomic overview of glioblastoma cells (A172) exposed to perillyl alcohol. *J. Proteomics*, **73**, 1018–1027.

DeLano,W. (2002) *PyMOL Molecular Graphics System*. PyMOL Schrödinger, http://www.pymol.org/.

Ferrari,A.J.R. *et al.* (2019) TopoLink: evaluation of structural models using chemical crosslinking distance constraints. *Bioinformatics*, **35**, 3169–3170.

Fischer,L. and Rappsilber,J. (2017) Quirks of error estimation in cross-linking/mass spectrometry. *Anal. Chem.*, **89**, 3829–3833.

Iglesias,A.H. *et al.* (2010) Identification of cross-linked peptides by high-resolution precursor ion scan. *Anal. Chem.*, **82**, 909–916.

Jung,T. *et al.* (2020) The polyglutamine expansion at the N-terminal of Huntingtin protein modulates the dynamic configuration and phosphorylation of the C-terminal HEAT domain. *Structure*, **28**, 1035–1050.e8.

Kahraman,A. *et al.* (2013) Cross-link guided molecular modeling with ROSETTA. *PLoS One*, **8**, e73411.

Lima,D.B. *et al.* (2018) Characterization of homodimer interfaces with cross-linking mass spectrometry and isotopically labeled proteins. *Nat. Protoc.*, **13**, 431–458.

Lloyd,S. (1982) Least squares quantization in PCM. *IEEE Trans. Inf. Theory*, **28**, 129–137.

Mintseris,J. and Gygi,S.P. (2020) High-density chemical cross-linking for modeling protein interactions. *Proc. Natl. Acad. Sci. USA*, **117**, 93–102.

Phillips,A.S. *et al.* (2015) Conformational dynamics of α-synuclein: insights from mass spectrometry. *Analyst*, **140**, 3070–3081.

Rousseeuw,P.J. (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.*, **20**, 53–65.

Sinz,A. *et al.* (2015) Chemical cross-linking and native mass spectrometry: a fruitful combination for structural biology: chemical crosslinking and native MS. *Protein Sci.*, **24**, 1193–1209.

Vizcaíno,J.A. *et al.* (2017) The mzIdentML data standard version 1.2, supporting advances in proteome informatics. *Mol. Cell. Proteomics*, **16**, 1275–1285.