Journal of Visualized Experiments

# Quaternary structure modeling through chemical cross-linking mass spectrometry: Extending TX-MS Jupyter reports.

--Manuscript Draft--

| | |
|---|---|
| **Article Type:** | Invited Methods Article - JoVE Produced Video |
| **Manuscript Number:** | JoVE60311R2 |
| **Full Title:** | Quaternary structure modeling through chemical cross-linking mass spectrometry: Extending TX-MS Jupyter reports. |
| **Section/Category:** | JoVE Biology |
| **Keywords:** | Protein-protein interactions<br>Host-pathogen interactions<br>Chemical Cross-linking Mass Spectrometry<br>Protein docking<br>Protein structure modeling<br>Jupyter Notebooks |
| **Corresponding Author:** | Lars Malmstroem<br>Institute for computational science<br>Zurich, Zurich SWITZERLAND |
| **Corresponding Author's Institution:** | Institute for computational science |
| **Corresponding Author E-Mail:** | lars.malmstroem@uzh.ch |
| **Order of Authors:** | Lars Malmstroem |
| | Hamed Khakzad |
| | Swen Vermeul |
| **Additional Information:** | |
| **Question** | **Response** |
| Please indicate whether this article will be Standard Access or Open Access. | Open Access (US$4,200) |
| Please indicate the **city, state/province, and country** where this article will be **filmed**. Please do not use abbreviations. | Zurich, Switzerland |

University of Zurich[UZH]

**Institute for Computational Science**

University of Zurich
Winterthurerstrasse 190
CH-8057 Zurich
www.ics.uzh.ch

**Lars Malmström**
Lars.malmstroem@uzh.ch

Zurich, 30 October 2019

Dear Editor Wu,

We are pleased to submit our revised manuscript, "*Quaternary structure modeling through chemical cross-linking mass spectrometry: Extending TX-MS Jupyter reports*", for consideration by JoVE. The manuscript describes a method to enable biologists with minimal technical skills to analyze structural models and the supporting MS data in more depth compared to the standard report produce by the TX-MS method which we recently published in Nature Communications[1]. All the changes to the manuscript were tracked and each point raised by reviewers addressed in a separate rebuttal letter. We have uploaded high-resolution versions of the figures. Given the easily addressed reviewer comments, we are cautiously optimistic that the manuscript will be accepted, and we are looking forward to the next steps.

Sincerely,

Dr. Lars Malmström
University of Zurich

1. Hauri, S. *et al.* Rapid determination of quaternary protein structures in complex biological samples. *Nat Commun* **10,** 192 (2019).

**TITLE:**

# Quaternary Structure Modeling Through Chemical Cross-Linking Mass Spectrometry: Extending TX-MS Jupyter Reports

**AUTHORS AND AFFILIATIONS:**

Hamed Khakzad[1,2], Swen Vermeul[3], Lars Malmström[1,2,4,5]

[1]Institute for Computational Science, University of Zurich, Zurich, Switzerland
[2]SIB Swiss Institute of Bioinformatics, Quartier Sorge, Lausanne, Switzerland
[3]ETH Zurich, Scientific IT Services, Zurich, Switzerland
[4]S3IT, University of Zurich, Zurich, Switzerland
[5]Division of Infection Medicine, Department of Clinical Sciences Lund, Faculty of Medicine, Lund University, Lund, Sweden

**Corresponding Author:**

Lars Malmström (lars.malmstroem@uzh.ch)

**Email Addresses of Co-authors:**

Hamed Khakzad (hamed.khakzad@uzh.ch)
Swen Vermeul (swen@ethz.ch)

**KEYWORDS:**

protein-protein interactions, host-pathogen interactions, chemical cross-linking mass spectrometry, protein docking, protein structure modeling, Jupyter Notebooks

**SUMMARY:**

Targeted cross-linking mass spectrometry creates quaternary protein structure models using mass spectrometry data acquired using up to three different acquisition protocols. When run as an iPortal workflow, the results are reported in a Jupyter Notebook. Here, we demonstrate how the Jupyter Notebook can be extended for a more in-depth analysis.

**ABSTRACT:**

Protein-protein interactions can be challenging to study yet provide insights into how biological systems function. Targeted cross-linking mass spectrometry (TX-MS), a method combining quaternary protein structure modeling and chemical cross-linking mass spectrometry, creates high-accuracy structure models using data obtained from complex, unfractionated samples. This removes one of the major obstacles to protein complex structure analysis, because the proteins of interest no longer need to be purified in large quantities. An automated computational workflow was developed to make the protocol more accessible to the community. The workflow creates a Jupyter Notebook, a graphical report summarizing the most important analysis results. Extending the Jupyter Notebook can yield more in-depth insights and a better understanding of the model and the mass spectrometry data supporting it. The protocol presented here demonstrates some of the most common extensions and explains what type of information can be obtained. It contains blocks to help analyze various MS acquisition data such as high-resolution

MS1 (hrMS1), data-dependent acquisition (DDA), and data-independent acquisition (DIA). The result of such analyses can be applied to structural models that are embedded in the notebook using NGLView.

**INTRODUCTION:**

Protein-protein interactions underpin the structure and function of biological systems. Having access to quaternary structures of proteins can provide insights into how two or more proteins interact to form high-order structures. Obtaining quaternary structures remains challenging: this is reflected in the comparatively small number of Protein DataBank (PDB) entries[1] containing more than one polypeptide. Protein-protein interactions can be studied with technologies such as X-ray crystallography, NMR, and cryo-EM, but obtaining a sufficient amount of purified protein under conditions where the methods can be applied can be time-consuming.

Chemical cross-linking mass spectrometry was developed to obtain experimental data on protein-protein interactions with fewer restrictions on sample preparation as mass spectrometry and can be used to acquired data on arbitrarily complex samples[2-9]. However, the combinatorial nature of the data analysis and the relatively small number of cross-linked peptides require that the samples be fractionated before analysis. To address this shortcoming, we developed TX-MS, a method that combines computational modeling with chemical cross-linking mass spectrometry[10]. TX-MS can be used on arbitrarily complex samples and is significantly more sensitive compared to previous methods[10]. It accomplishes this by scoring all data associated with a given protein-protein interaction as a set, instead of interpreting each MS spectrum independently. TX-MS also uses up to three different MS acquisition protocols: high-resolution MS1 (hrMS1), data-dependent acquisition (DDA) MS, and data-independent acquisition (DIA) MS), further providing opportunities to identify a cross-linked peptide by combining multiple observations. The TX-MS computational workflow is complex for several reasons. First, it relies on multiple MS analysis software programs[11-13] to create protein structure models[14,15]. Second, the amount of data can be large. Third, the modeling step can consume significant amounts of computer processing power. Consequently, TX-MS is best used as an automated computational iPortal workflow[16] that runs on large computational infrastructures such as computer clouds or clusters. To facilitate the interpretation of the results, we produced an interactive Jupyter Notebook[17]. Here, we demonstrate how the Jupyter Notebook report can be extended to yield a more in-depth analysis of a given result.

**PROTOCOL:**

**1. Register for an account at** https://txms.org.

1.1. Go to https://txms.org and click "**Sign Up**".

NOTE: We use the Rosetta modeling software. Because its usage is restricted to commercial users, we require registration to verify that the user belongs to an academic institution or a nonprofit organization.

1.2. Provide the requested information (i.e., name and email). Wait for the confirmation email with the user credentials.

NOTE: It can take up to 24 h to receive the user credentials because we need to manually verify that the user is from an academic institution.

**2. Run TX-MS.**

NOTE: Convert the vendor specific formats to mzML using the ProteoWizard MSConvert software[19].

2.1. Upload the MS data to https://txms.org. Click on "**Upload Data**" and pull the MS data, which must be in the mzML data format[18], to the upload box.

NOTE: Example data are available on https://txms.org. These data are also directly accessible through zenodo.org, DOI 10.5281/zenodo.3361621.

2.2. Upload two PDB files to https://txms.org. Click on "**Upload Data**" and pull the PDB files to the upload box.

NOTE: If no experimental structures exist, create models using, for example, Robetta[20,21].

2.3. Submit a new workflow and select the uploaded MS data and PDB files. Click on "**Submit Workflow**", and type in a title and a description. Click "**View Workflow**" and select the "**Cheetah**" workflow. Follow the wizard.

NOTE: Computing the result takes time, so please wait until the workflow finishes. The computation is carried out on a remote computational infrastructure. If you want to run TX-MS locally, please refer to Hauri et al.[10].

2.4. Inspect the Jupyter Notebook report using the online viewer. Click on "**View Workflow**" and scroll down to "**Report**".

**3. Install JupyterHub.**

3.1. Install docker as instructed at https://docs.docker.com/install/.

3.2. Download the JupyterHub docker container with the Jupyter openBIS extension. The general command is "**docker pull malmstroem/jove:latest**", but might differ on other platforms.

NOTE: For a general discussion on how to download containers, please refer to https://www.docker.com/get-started. It is also possible to download the container from zenodo.org, DOI 10.5281/zenodo.3361621.

NOTE: The Jupyter openBIS extension source code is available here: https://pypi.org/project/jupyter-openbis-extension/.

3.3. Start the docker container: docker run -p 8178:8000 malmstroem/jove:latest.

NOTE: The port that JupyterHub uses by default is 8000. This port is configurable, and the commands above need to be adjusted accordingly if changed. Port 8178 is an arbitrary choice and can be changed. The example URLs provided below need to be adjusted accordingly.

3.4. Go to the following address: http://127.0.0.1:8178. Log in using the username "**user**" and the password "**user**".

NOTE: The address http://127.0.0.1 implies that the docker container is running on the local computer. If the docker container is run on a server, use the server's IP address or URL (e.g., https://example.com). The docker container is based on Ubuntu Bionic 18.04, JupyterHub 0.9.6, and Jupyter openBIS extension 0.2. It is possible to install this in other operating systems, but this was not tested.

**4. Download the report.**

4.1. Create a new notebook by clicking **New**|**Python3** using the menu located near the top right part of the page. This will open a new tab with a notebook called **Untitled** (or something similar).

4.2. Click "**Configure openBIS Connections**" in the Jupyter tool menu.

4.3. Fill in the name: txms; url: https://txms.org; user: guest; password: guestpasswd.

4.4. Click "**Connect**".

4.5. Choose the new connection and click "**Choose Connection**".

4.6. Search for the report and click **Download**.

NOTE: There are many ways to search for the result. The easiest is to type in the workflow code (e.g., /CHEETAH/WF70).

4.7. Rerun the report by clicking **Cell**|**Run All**.

**5. Extend the report.**

5.1. Add a new cell at the bottom: **Cell**|**Insert Below**.

5.2. Type in the wanted code. For an example, please see the Representative Results section below.

5.3. Execute the cell by pressing "**Shift-Enter**".

**6. Upload the new report.**

6.1. Create a new dataset by clicking the **Upload** button.

6.2. Select the new report and upload it.

**REPRESENTATIVE RESULTS:**

TX-MS provides structural outputs supported by MS-derived experimental constraints. It works by combining different MS data acquisition types with computational modeling. It is useful to parse each MS data separately and provide visualization of the output structure. **Supplementary Data 1** contains an example notebook that can parse DDA and DIA data produced as the output of TX-MS. Users can select the XL of interest and, by running the notebook, the MS2 spectrum of that XL will be shown where different colors help to discriminate between fragments related to the first peptide, second peptide, and the combination. The XL can also be mapped to the structure using the NGLView widget embedded in a Jupyter Notebook.

Another cell in this notebook can help users to parse and visualize DIA data. Visualizing DIA data is more difficult, because the analyzed data need to be prepared in the correct format.

**Figure 1** shows an example structure of M1 and albumin with top XLs mapped on the structure. All XLs were obtained by TX-MS after parsing hrMS1, DDA, and DIA data, and the computational models were provided by the RosettaDock protocol.

As this report is a Jupyter Notebook, any valid Python code can be added to new notebook cells. For example, the code below will create a histogram over the MS2 counts, an indication for how well supported each cross-link is by the underlying data.

```
import seaborn as sns
sns.distplot(ms2['count']);
```

**FIGURE AND TABLE LEGENDS:**

**Figure 1: Structural model of *Streptococcus pyogenes* protein M1 and human albumin with XLs mapped on the structure.** The M protein is shown in gray and constitutes a homodimer. The six albumin molecules are presented as pairs in various shades of blue. Cross-links and distances are presented in red with black text.

**Supplementary File. Jupyter notebook data.**

**DISCUSSION:**

Modern computational workflows are often complex with multiple tools from many different vendors and complex interdependencies, high data volumes, and multifaceted results. Consequently, it is increasingly difficult to accurately document all the steps required to obtain a

result, thereby making it difficult to reproduce the given result. Here, we demonstrate a general strategy that combines the automation and ease of an automated workflow that produces a generic report, with the flexibility to customize the report in a reproducible fashion.

Three requirements need to be fulfilled for the protocol to work: First, the proteins selected for analysis need to interact in such a manner that the chemical cross-linking experiment can produce cross-linked species at a sufficiently high concentration to be detected by the mass spectrometer; different mass spectrometers have different levels of detection and are also dependent on the acquisition protocol as well as the choice of cross-linking reagent. The TX-MS protocol only allows for DSS, a lysine-lysine homobifunctional cross-linking reagent, but this limitation is largely due to the possibility that the machine learning step would need to be adjusted for other reagents. Second, the two proteins need either to have an experimentally determined structure, or need to be modeled using either comparative modeling techniques or de novo techniques. Not all proteins can be modeled, but a combination of improved software, as well as a constant deposition of experimental structures in the PDB expands the number of proteins that can be modeled. Third, the interacting proteins should remain sufficiently similar in their bound and unbound states so that the flexible backbone-docking algorithms in use by TX-MS can create quaternary structures of adequate quality to enable scoring. This requirement is relatively vague, as adequate quality is highly system-dependent, where smaller proteins of known structure are generally easier to compare than larger proteins of unknown structure.

In case of a negative result, first check that TX-MS found intra-links, that is, cross-links between residues that are part of the same polypeptide chain. If none are discovered, the most likely explanation is that something went wrong with the sample preparation or the data acquisition. If multiple distance constraints do not support the models, visually inspect the models to ensure that the conformation is supported by cross-linked residues in such a way that there is no obvious way to pivot one of the interactors without disrupting at least one cross-link. If there are cross-links longer than the permitted distance for the given cross-linking reagent, try to improve the modeling of the interactors by incorporating cross-linking data.

It is possible to use alternative software applications to accomplish equivalent results provided that the sensitivity of the chosen software is equivalent to the sensitivity of TX-MS. For example, there are online versions of RosettaDock, HADDOCK, and others. It is also possible to analyze chemical cross-linking data through xQuest/xProphet[5,6], plink[7], and SIM-XL[23].

We are continuously applying TX-MS to new projects[24] and thereby improving the reports produced by TX-MS to allow for a more detailed analysis of results without having to make the reports larger.

**ACKNOWLEDGMENTS:**
This work was supported by Foundation of Knut and Alice Wallenberg (grant no. 2016.0023) and the Swiss National Science Foundation (grant no. SNF 200021 160188). We thank S3IT, University of Zurich for computational infrastructure and technical support.

**DISCLOSURES:**

The authors have nothing to disclose.

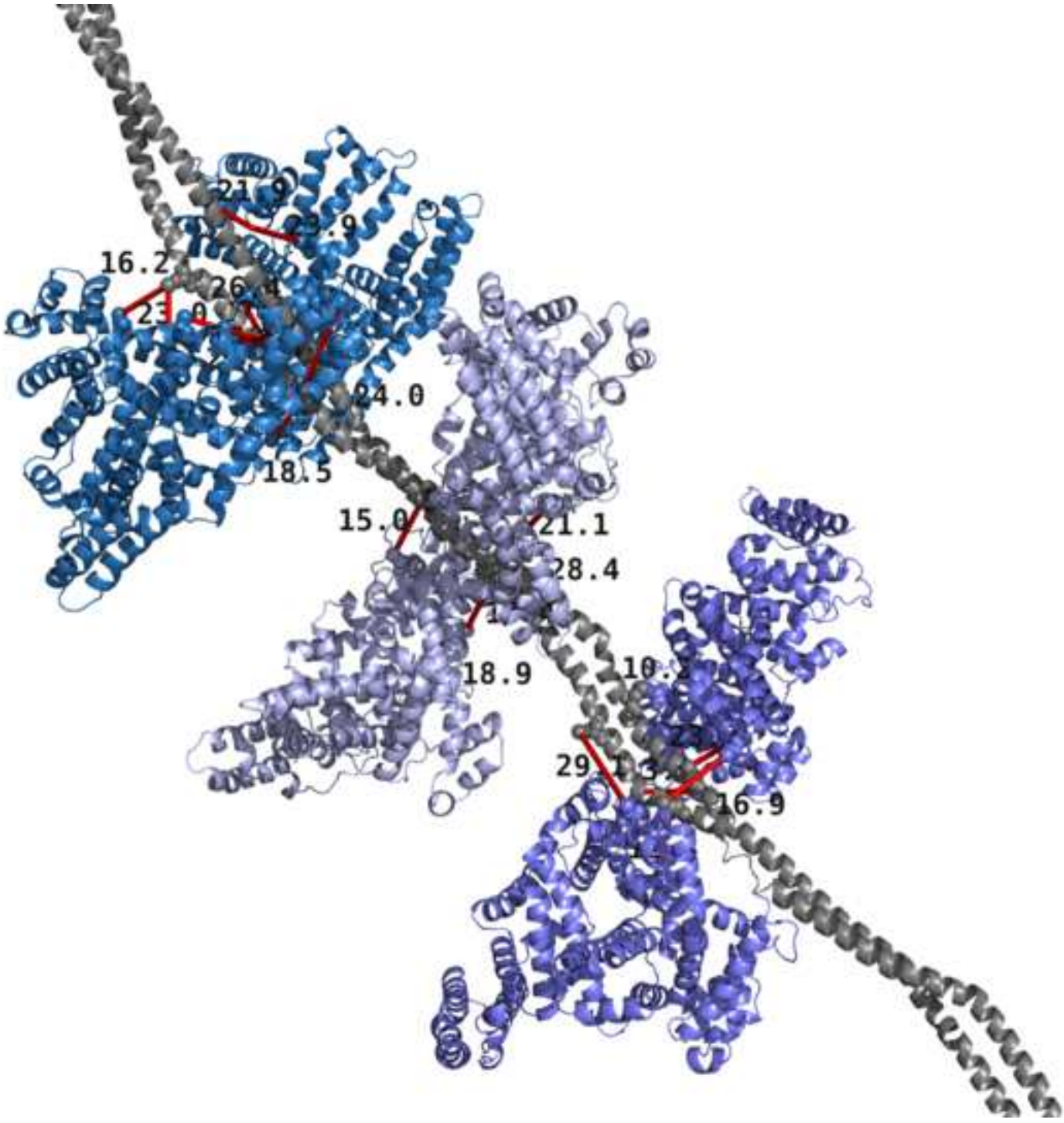
**REFERENCES:**


1. Berman, H. M. et al. The Protein Data Bank. *Acta Crystallographica Section D: Biological Crystallography*. **58** (6), 899–907 (2002).
2. Herzog, F., et al. Structural Probing of a Protein Phosphatase 2A Network by Chemical Cross-Linking and Mass Spectrometry. *Science*. **337** (6100), 1348–1352 (2012).
3. Hoopmann, M. R. et al. Kojak: efficient analysis of chemically cross-linked protein complexes. *Journal of Proteome Research*. **14** (5), 2190–2198 (2015).
4. Seebacher, J. et al. Protein cross-linking analysis using mass spectrometry, isotope-coded cross-linkers, and integrated computational data processing. *Journal of Proteome Research*. **5** (9), 2270–2282 (2006).
5. Rinner, O. et al. Identification of cross-linked peptides from large sequence databases. *Nature Methods*. **5** (4), 315–318 (2008).
6. Walzthoeni, T. et al. False discovery rate estimation for cross-linked peptides identified by mass spectrometry. *Nature Methods*. **9** (9), 901–903 (2012).
7. Yang, B. et al. Identification of cross-linked peptides from complex samples. *Nature Methods*. **9** (9), 904–906 (2012).
8. Chu, F., Baker, P. R., Burlingame, A. L., Chalkley, R. J. Finding Chimeras: a Bioinformatics Strategy for Identification of Cross-linked Peptides. *Molecular & Cellular Proteomics*. **9** (1), 25–31 (2010).
9. 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. *Journal of Proteome Research*. **12** (12), 5923–5933 (2013).
10. Hauri, S. et al. Rapid determination of quaternary protein structures in complex biological samples. *Nature Communications*. **10** (1), 192 (2019).
11. Röst, H. L. et al. OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. *Nature Biotechnology*. **32** (3), 219–223 (2014).
12. Röst, H. L. et al. OpenMS: a flexible open-source software platform for mass spectrometry data analysis. *Nature Methods.* **13** (9), 741–748 (2016).
13. Quandt, A. et al. Using synthetic peptides to benchmark peptide identification software and search parameters for MS/MS data analysis. *EuPA Open Proteomics*. **5**, 21–31 (2014).
14. Bradley, P. et al. Free modeling with Rosetta in CASP6. *Proteins: Structure, Function, and Bioinformatics*. **61** (S7), 128–134 (2005).
15. Gray, J. J. High-resolution protein–protein docking. *Current Opinion in Structural Biology*. **16** (2), 183–193 (2006).
16. Kunszt, P. et al. iPortal: the swiss grid proteomics portal: Requirements and new features based on experience and usability considerations. *Concurrency and Computation: Practice and Experience*. **27** (2), 433–445 (2014).
17. Malmström, L. Computational Proteomics with Jupyter and Python. *Methods in Molecular Biology*. *(Clifton, N.J.)* **1977** (Chapter 15), 237–248 (2019).
18. Martens, L. et al. mzML--a community standard for mass spectrometry data. *Molecular & Cellular Proteomics*. **10** (1), R110.000133 (2011).

19. Chambers, M. C. et al. A cross-platform toolkit for mass spectrometry and proteomics. *Nature Biotechnology*. **30** (10), 918–920 (2012).
20. Chivian, D. et al. Prediction of CASP6 structures using automated Robetta protocols. *Proteins: Structure, Function, and Bioinformatics*. **61 Suppl 7** (S7), 157–166 (2005).
21. Chivian, D. et al. Automated prediction of CASP-5 structures using the Robetta server. *Proteins: Structure, Function, and Bioinformatics*. **53** (S6), 524–533 (2003).
22. Bauch, A. et al. openBIS: a flexible framework for managing and analyzing complex data in biology research. *BMC Bioinformatics*. **12**, 468 (2011).
23. Lima, D. B. et al. SIM-XL: A powerful and user-friendly tool for peptide cross-linking analysis. *Journal of Proteomics*. **129**, 51–55 (2015).
24. Happonen, L. et al. A quantitative Streptococcus pyogenes–human protein–protein interaction map reveals localization of opsonizing antibodies. *Nature Communications*. **10**, 2727 (2019).

Figure

Click here to access/download;Figure;figure1.png



| Name of Material/Equipment | Company |
| --- | --- |
| Two Protein DataBank files of the proteins of interest. | N/A |
| An mzML data file acquired on a sample where the proteins of interest were crosslinked. | N/A |

| Catalog Number | Comments/Description |
|---|---|
| **N/A** | **Example files available on txms.org and zenodo.org, DOI 10.5281/zenodo.3361621** |
| **N/A** | **Example files available on txms.org or zenodo.org, DOI 10.5281/zenodo.3361621** |

jove JOURNAL OF VISUALIZED EXPERIMENTS
1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com

# ARTICLE AND VIDEO LICENSE AGREEMENT

Title of Article: Extending TxMS reports using Dugutu Method

Author(s): Hamed Khakzad, Sven Vermeul, Lars Malmström

Item 1: The Author elects to have the Materials be made available (as described at http://www.jove.com/publish) via:

☐ Standard Access ☒ Open Access

Item 2: Please select one of the following items:

☒ The Author is **NOT** a United States government employee.

☐ The Author is a United States government employee and the Materials were prepared in the course of his or her duties as a United States government employee.

☐ The Author is a United States government employee but the Materials were NOT prepared in the course of his or her duties as a United States government employee.

## ARTICLE AND VIDEO LICENSE AGREEMENT

1. **Defined Terms.** As used in this Article and Video License Agreement, the following terms shall have the following meanings: "**Agreement**" means this Article and Video License Agreement; "**Article**" means the article specified on the last page of this Agreement, including any associated materials such as texts, figures, tables, artwork, abstracts, or summaries contained therein; "**Author**" means the author who is a signatory to this Agreement; "**Collective Work**" means a work, such as a periodical issue, anthology or encyclopedia, in which the Materials in their entirety in unmodified form, along with a number of other contributions, constituting separate and independent works in themselves, are assembled into a collective whole; "**CRC License**" means the Creative Commons Attribution-Non Commercial-No Derivs 3.0 Unported Agreement, the terms and conditions of which can be found at: http://creativecommons.org/licenses/by-nc-nd/3.0/legalcode; "**Derivative Work**" means a work based upon the Materials or upon the Materials and other pre-existing works, such as a translation, musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which the Materials may be recast, transformed, or adapted; "**Institution**" means the institution, listed on the last page of this Agreement, by which the Author was employed at the time of the creation of the Materials; "**JoVE**" means MyJove Corporation, a Massachusetts corporation and the publisher of The Journal of Visualized Experiments; "**Materials**" means the Article and / or the Video; "**Parties**" means the Author and JoVE; "**Video**" means any video(s) made by the Author, alone or in conjunction with any other parties, or by JoVE or its affiliates or agents, individually or in collaboration with the Author or any other parties, incorporating all or any portion of the Article, and in which the Author may or may not appear.

2. **Background.** The Author, who is the author of the Article, in order to ensure the dissemination and protection of the Article, desires to have the JoVE publish the Article and create and transmit videos based on the Article. In furtherance of such goals, the Parties desire to memorialize in this Agreement the respective rights of each Party in and to the Article and the Video.

3. **Grant of Rights in Article.** In consideration of JoVE agreeing to publish the Article, the Author hereby grants to JoVE, subject to **Sections 4** and **7** below, the exclusive, royalty-free, perpetual (for the full term of copyright in the Article, including any extensions thereto) license (a) to publish, reproduce, distribute, display and store the Article in all forms, formats and media whether now known or hereafter developed (including without limitation in print, digital and electronic form) throughout the world, (b) to translate the Article into other languages, create adaptations, summaries or extracts of the Article or other Derivative Works (including, without limitation, the Video) or Collective Works based on all or any portion of the Article and exercise all of the rights set forth in (a) above in such translations, adaptations, summaries, extracts, Derivative Works or Collective Works and(c) to license others to do any or all of the above. The foregoing rights may be exercised in all media and formats, whether now known or hereafter devised, and include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. If the "Open Access" box has been checked in **Item 1** above, JoVE and the Author hereby grant to the public all such rights in the Article as provided in, but subject to all limitations and requirements set forth in, the CRC License.

4. **Retention of Rights in Article.** Notwithstanding the exclusive license granted to JoVE in **Section 3** above, the Author shall, with respect to the Article, retain the non-exclusive right to use all or part of the Article for the non-commercial purpose of giving lectures, presentations or teaching classes, and to post a copy of the Article on the Institution's website or the Author's personal website, in each case provided that a link to the Article on the JoVE website is provided and notice of JoVE's copyright in the Article is included. All non-copyright intellectual property rights in and to the Article, such as patent rights, shall remain with the Author.

5. **Grant of Rights in Video – Standard Access.** This **Section 5** applies if the "Standard Access" box has been checked in **Item 1** above or if no box has been checked in **Item 1** above. In consideration of JoVE agreeing to produce, display or otherwise assist with the Video, the Author hereby acknowledges and agrees that, Subject to **Section 7** below, JoVE is and shall be the sole and exclusive owner of all rights of any nature, including, without limitation, all copyrights, in and to the Video. To the extent that, by law, the Author is deemed, now or at any time in the future, to have any rights of any nature in or to the Video, the Author hereby disclaims all such rights and transfers all such rights to JoVE.

6. **Grant of Rights in Video – Open Access.** This **Section 6** applies only if the "Open Access" box has been checked in **Item 1** above. In consideration of JoVE agreeing to produce, display or otherwise assist with the Video, the Author hereby grants to JoVE, subject to **Section 7** below, the exclusive, royalty-free, perpetual (for the full term of copyright in the Article, including any extensions thereto) license (a) to publish, reproduce, distribute, display and store the Video in all forms, formats and media whether now known or hereafter developed (including without limitation in print, digital and electronic form) throughout the world, (b) to translate the Video into other languages, create adaptations, summaries or extracts of the Video or other Derivative Works or Collective Works based on all or any portion of the Video and exercise all of the rights set forth in (a) above in such translations, adaptations, summaries, extracts, Derivative Works or Collective Works and (c) to license others to do any or all of the above. The foregoing rights may be exercised in all media and formats, whether now known or hereafter devised, and include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. For any Video to which this **Section 6** is applicable, JoVE and the Author hereby grant to the public all such rights in the Video as provided in, but subject to all limitations and requirements set forth in, the CRC License.

7. **Government Employees.** If the Author is a United States government employee and the Article was prepared in the course of his or her duties as a United States government employee, as indicated in **Item 2** above, and any of the licenses or grants granted by the Author hereunder exceed the scope of the 17 U.S.C. 403, then the rights granted hereunder shall be limited to the maximum rights permitted under such statute. In such case, all provisions contained herein that are not in conflict with such statute shall remain in full force and effect, and all provisions contained herein that do so conflict shall be deemed to be amended so as to provide to JoVE the maximum rights permissible within such statute.

8. **Protection of the Work.** The Author(s) authorize JoVE to take steps in the Author(s) name and on their behalf if JoVE believes some third party could be infringing or might infringe the copyright of either the Author's Article and/or Video.

9. **Likeness, Privacy, Personality.** The Author hereby grants JoVE the right to use the Author's name, voice, likeness, picture, photograph, image, biography and performance in any way, commercial or otherwise, in connection with the Materials and the sale, promotion and distribution thereof. The Author hereby waives any and all rights he or she may have, relating to his or her appearance in the Video or otherwise relating to the Materials, under all applicable privacy, likeness, personality or similar laws.

10. **Author Warranties.** The Author represents and warrants that the Article is original, that it has not been published, that the copyright interest is owned by the Author (or, if more than one author is listed at the beginning of this Agreement, by such authors collectively) and has not been assigned, licensed, or otherwise transferred to any other party. The Author represents and warrants that the author(s) listed at the top of this Agreement are the only authors of the Materials. If more than one author is listed at the top of this Agreement and if any such author has not entered into a separate Article and Video License Agreement with JoVE relating to the Materials, the Author represents and warrants that the Author has been authorized by each of the other such authors to execute this Agreement on his or her behalf and to bind him or her with respect to the terms of this Agreement as if each of them had been a party hereto as an Author. The Author warrants that the use, reproduction, distribution, public or private performance or display, and/or modification of all or any portion of the Materials does not and will not violate, infringe and/or misappropriate the patent, trademark, intellectual property or other rights of any third party. The Author represents and warrants that it has and will continue to comply with all government, institutional and other regulations, including, without limitation all institutional, laboratory, hospital, ethical, human and animal treatment, privacy, and all other rules, regulations, laws, procedures or guidelines, applicable to the Materials, and that all research involving human and animal subjects has been approved by the Author's relevant institutional review board.

11. **JoVE Discretion.** If the Author requests the assistance of JoVE in producing the Video in the Author's facility, the Author shall ensure that the presence of JoVE employees, agents or independent contractors is in accordance with the relevant regulations of the Author's institution. If more than one author is listed at the beginning of this Agreement, JoVE may, in its sole

jove VISUALIZED EXPERIMENTS
1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com

# ARTICLE AND VIDEO LICENSE AGREEMENT

discretion, elect not take any action with respect to the Article until such time as it has received complete, executed Article and Video License Agreements from each such author. JoVE reserves the right, in its absolute and sole discretion and without giving any reason therefore, to accept or decline any work submitted to JoVE. JoVE and its employees, agents and independent contractors shall have full, unfettered access to the facilities of the Author or of the Author's institution as necessary to make the Video, whether actually published or not. JoVE has sole discretion as to the method of making and publishing the Materials, including, without limitation, to all decisions regarding editing, lighting, filming, timing of publication, if any, length, quality, content and the like.

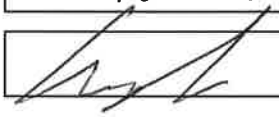
12. **Indemnification.** The Author agrees to indemnify JoVE and/or its successors and assigns from and against any and all claims, costs, and expenses, including attorney's fees, arising out of any breach of any warranty or other representations contained herein. The Author further agrees to indemnify and hold harmless JoVE from and against any and all claims, costs, and expenses, including attorney's fees, resulting from the breach by the Author of any representation or warranty contained herein or from allegations or instances of violation of intellectual property rights, damage to the Author's or the Author's institution's facilities, fraud, libel, defamation, research, equipment, experiments, property damage, personal injury, violations of institutional, laboratory, hospital, ethical, human and animal treatment, privacy or other rules, regulations, laws, procedures or guidelines, liabilities and other losses or damages related in any way to the submission of work to JoVE, making of videos by JoVE, or publication in JoVE or elsewhere by JoVE. The Author shall be responsible for, and shall hold JoVE harmless from, damages caused by lack of sterilization, lack of cleanliness or by contamination due to the making of a video by JoVE its employees, agents or independent contractors. All sterilization, cleanliness or decontamination procedures shall be solely the responsibility of the Author and shall be undertaken at the Author's expense. All indemnifications provided herein shall include JoVE's attorney's fees and costs related to said losses or damages. Such indemnification and holding harmless shall include such losses or damages incurred by, or in connection with, acts or omissions of JoVE, its employees, agents or independent contractors.

13. **Fees.** To cover the cost incurred for publication, JoVE must receive payment before production and publication of the Materials. Payment is due in 21 days of invoice. Should the Materials not be published due to an editorial or production decision, these funds will be returned to the Author. Withdrawal by the Author of any submitted Materials after final peer review approval will result in a US$1,200 fee to cover pre-production expenses incurred by JoVE. If payment is not received by the completion of filming, production and publication of the Materials will be suspended until payment is received.

14. **Transfer, Governing Law.** This Agreement may be assigned by JoVE and shall inure to the benefits of any of JoVE's successors and assignees. This Agreement shall be governed and construed by the internal laws of the Commonwealth of Massachusetts without giving effect to any conflict of law provision thereunder. This Agreement may be executed in counterparts, each of which shall be deemed an original, but all of which together shall be deemed to me one and the same agreement. A signed copy of this Agreement delivered by facsimile, e-mail or other means of electronic transmission shall be deemed to have the same legal effect as delivery of an original signed copy of this Agreement.

A signed copy of this document must be sent with all new submissions. Only one Agreement is required per submission.

**CORRESPONDING AUTHOR**

| | |
|---|---|
| Name: | Lars Malmström |
| Department: | Institute for Computational Science |
| Institution: | University of Zurich |
| Title: | Senior Scientist |
| Signature: |  |
| Date: | 24 May 2019 |

Please submit a **signed** and **dated** copy of this license by one of the following three methods:

1. Upload an electronic version on the JoVE submission site
2. Fax the document to +1.866.381.2236
3. Mail the document to JoVE / Attn: JoVE Editorial / 1 Alewife Center #200 / Cambridge, MA 02140

Rebuttal letter

We are pleased for the opportunity to submit a revised manuscript and appreciate the reviewer's and editorial staff's comments and insights. We were able to address all the concerns and hope that the reviewers and editors find the manuscript suitable for publication in the Journal of Visualized Experiments.

**Editorial comments:**

1. Please take this opportunity to thoroughly proofread the manuscript to ensure that there are no spelling or grammar issues.

We have gone over the text in detail and have removed spelling and grammatical issues.

2. Please obtain explicit copyright permission to reuse any figures (including TOC) from a previous publication. Explicit permission can be expressed in the form of a letter from the editor or a link to the editorial policy that allows re-prints. Please upload this information as a .doc or .docx file to your Editorial Manager account. The Figure must be cited appropriately in the Figure Legend, i.e. "This figure has been modified from [citation]."

All the figures were newly created for this paper; data from previous papers were re-used, but the analysis was done specifically for this paper. We do not believe that we need to ask for permission to re-use the data as long as we cite the original paper. In addition, the previous papers were published in Nature Communication, a journal under a generous Creative Commons license stating:
"Under this license users are free to share (copy, distribute and transmit) and remix (adapt) the contribution including for commercial purposes, providing they attribute the contribution in the manner specified by the author or licensor"

and

"Under Creative Commons, authors retain copyright in their articles."

For more details:
https://www.nature.com/ncomms/about/open-access

As such, we have fulfilled the license requirements by citing the paper.

3. Please do not use abbreviation in the title.

TX-MS is not an abbreviation but the name of a specific method. This name is an acronym, but the issue here is that we do specifically refer to the TX-MS method, not to "targeted cross-linking mass spectrometry" in general. We have, however, updated the title in accordance with a reviewer comment and we hope that the new title is permitted.

4. Please ensure that the references appear as the following:
Lastname, F.I., LastName, F.I., LastName, F.I. Article Title. Source. Volume (Issue), FirstPage – LastPage, (YEAR).
For more than 6 authors, list only the first author then et al.

We are adhering to the requested citation format.

5. Step 4.2: Please specify the code.

The idea of the paper is that this input comes from the user, to provide the opportunity for the user to explore the data in greater detail compared to the original report.

6. Please remove the embedded figure(s) from the manuscript. All figures should be uploaded separately to your Editorial Manager account.

We have removed the embedded figures and uploaded the figures using the editorial manager account.

7. Each figure must be accompanied by a title and a description after the Representative Results of the manuscript text.

We have accompanied each figure with a title and description.

8. Please revise the Discussion to explicitly cover the following in detail in 3-6 paragraphs with citations:
a) Critical steps within the protocol
b) Any modifications and troubleshooting of the technique
c) Any limitations of the technique
d) The significance with respect to existing methods
e) Any future applications of the technique

We have extended the discussion to cover the request explicitly.

9. Please revise the table of materials to include all essential supplies, reagents, and equipment. The table should include the name, company, and catalog number of all relevant materials in separate columns in an xls/xlsx file.

We have revised the materials.

**Reviewers' comments:**
**Reviewer #1:**
The manuscript as presented is minimal and it was not possible to fully review the work. What could be reviewed was not substantial enough to provide significant help to a non-expert and was poor for an expert. In general, the protocol only provides minimal detail and should be substantially expanded. Few of the steps are clear and the data linked from the manuscript is not easily found.

We have extended the manuscript to provide more help for users. We have also uploaded the needed input files and containers to zenodo.org and provide the DOI (10.5281/zenodo.3361621 ) in the manuscript.

Specific points are found below, these are not by any means exhaustive.
1) It feels highly disingenuous that they appear not to have cited any previous XLMS software?
E.g. xQuest (Rinner et al., Nat Methods 5(4):315-8, 2008; Walzthoeni et al., Nat Methods 9(9):901-3, 2012), pLink (Yang et al., Nat Methods 9(9):904-6, 2012), ProteinProspector (Chu et al., Mol Cell Proteomics 9:25-31, 2010; Trnka et al., 13(2):420-34, 2014), Hekate (Holding et al. J Proteome Res 12(12) 5923-593, 2013) and Kojak (Hoopmann et al., J Proteome Res 14(5):2190-198, 2015) etc.

We are now citing the papers pointed out by the reviewer.

2) On line 69 they mention that this method is "significantly more sensitive compared to previous methods" highlighting the lack of references to previous methods and without substantiating this claim.

We are referring to the TX-MS paper where we did indeed substantiate this claim to the satisfaction of the reviewers of that paper. The original submission of that paper contained a supplementary note that detailed our performance compared to established tools; however, that supplementary note was left out from the final paper as, through discussions with the Nature Communication editorial staff, we felt that such a comparison could not be carried out objectively as we know our tools better than the established tools. The supplementary note, in short, demonstrated that TX-MS was the only tool able to identify inter-crosslinks between proteins in the most complex samples, human plasma adsorbed onto live bacteria.

3) I cannot properly review step 1 - Run TX-MS as this requires me to sign up with name and organisation. As this is under peer review and to sign up would bypass the anonymous/blind peer nature of this review.

This is rather unfortunate, we sent login credentials for five reviewer accounts to JoVE to be sent to the reviewers, but something must have gone wrong. The login accounts are:

reviewer1 VaayMM1D
reviewer2 XM2CCipW
reviewer3 5rpu4TBi
reviewer4 8UHrPWzq
reviewer5 WG29f6yv

4) On line 94 it states "NOTE: There is example data available on txms.org", this should be included as a supplementary file or directly linked with a doi.

We uploaded data and containers to zenodo.org under DOI 10.5281/zenodo.3361621. The data from the TX-MS paper is also on zenodo.org and can be used as input, under DOI 10.5281/zenodo.3361621.

5) On line 96 they mention uploading two structures - this implies the method is limited to this number. The authors should clarify.

The current implementation is limited to two; we have clarified this in the text.

6) On line 101 it states "NOTE: There is example data available on txms.org" this should be included as a supplementary file or directly linked with a doi.

We uploaded data and containers to zenodo.org under DOI 10.5281/zenodo.3361621. The data from the TX-MS paper is also on zenodo.org and can be used as input, under DOI 10.5281/zenodo.3361621.

7) On line 115 it is not clear if "docker pull 116 malmstroem/jove:latest" a Windows, Mac or Linux terminal command. How is the average scientist meant to realise this? I had to run it in the "Docker Quickstart Terminal" not on the mac terminal.

We have made the instructions clearer by providing instructions for each operating system.

8) On line "Start the docker container: docker run -p 8178:8080 malmstroem/jove:latest" I have to correct 8080 to 8000 to make this command work.

We apologize for this error and have corrected it.

9) "Click on the log-in button on the openBIS Jupyter extension" Jupyter asked for a username/password to "Sign-in" that I have not been provided. (I found " guest/guestpasswd" and "user/user" on docker hub, but why I am searching for these?).

The fundamental issue is that we do require to log in due to the licensing for Rosetta. Rosetta is only free to use by academics, and hence, we need to verify that users are indeed academic. We have made changes to the manuscript clarifying this.

10) On line "Search for the report and click download" - is completely unclear how to do

this. Again I worked it out from the dockerhub page, but that wasn't particularly helpful either. At this point, I stopped as too much required information outside of the manuscript.

We have improved the manuscript to include better instructions.

11) On line 136 is states "We provide a few examples under the Representative results section" - these are not clearly signposted.

We have clarified this statement.

12) Table of Materials - "Two Protein DataBank files of the proteins of interest." - These should be either a supplementary file, linked to a copy with a DOI, or a direct link to PDB. As it currently stands it is an exercise for the reader to work out where the files are.

We uploaded data and containers to zenodo.org under DOI 10.5281/zenodo.3361621. One of these structures is a model and hence is not present in the PDB.

13) Table of Materials - "An mzML data file acquired on a sample where the proteins of interest were crosslinked/Example files available on txms.org." - should be placed with a link to the DOI file or attached directly to the manuscript, I could not find them on the website.

We uploaded data and containers to zenodo.org under DOI 10.5281/zenodo.3361621.

**Reviewer #2:**

Manuscript Summary:

Studying Protein- Protein Interactions (PPIs) has become very important for understanding biological phenomenon, esp. the origin of diseases. TX-MS, discussed in this manuscript/ video, is a new method for studying PPIs using Jupyter notebook. It combines Chemical Crossslinking (CXL)-Mass spectrometry (MS) with docking methods - bioinformatics tools (usually developed according to each laboratory's special requirements and writing the required Algorithms). Is this method dependant on Linux systems alone? Would that constitute a limitation? Adding docking data and analyzing these is an added step. We have also done such attempts. Will these necessarily add more quality to the results? Refining cryo EM results using CXL data is still time consuming even if one is using sophisticated software like HADDOCK (High Ambiguity Driven biomolecular DOCKing). TX-MS approach uses "Various MS acquisition data such as hrMS1, DDA, and DIA using the notebook using the NGL view".

We appreciate the reviewer's comments. This method is not dependent on Linux per se, but installation on other operating systems is challenging. Docker is solving this problem, providing a way to run the service anywhere. In this particular case, we carry out all the calculations as a service, allowing anyone with a browser to run TX-MS.

The example given describes only use of a Homobifunctionl Crosslinker, that too, only a di NHS crosslinker, which target mainly Lysine r residues and are less useful for studying intermolecular interactions. While it is true that Lysine residues are in abundance in proteins, yet PPI's are not restricted to studying Lysines alone. In the manuscript there is no mention of Heterobifunctional crosslinkers. Reference to A Sinz's work is completely missing-Pl do cross check this statement of mine. No reference is made to StavroX or MeroX (the latter being more suited for MS/MS cleavable crosslinkers), software developed in her laboratory, which help in identifying Intermolecular cross-linking better. For this StavroX/ MeroX software come in handy. Identifying the two crosslinked peptides/ proteins fragments is very difficult and usually only one of these is actually identified. In our recent paper in 2019, we have successfully identified the crosslinked fragments involving two proteins. Does TX-MS have the capability of doing so. For Chemical Crosslinking-Mass Spectrometry- Bioinformatics to emerge as a rapid, routine and reliable technique in hospital situations with large number of patient samples still remains a distant dream indeed.

The current implementation is only for a small subset of cross-linkers. The reason for this is that we have trained the machine learning step on this type of data; it is possible that the model also works for other cross-linkers, but we would like to verify this before making any claims. All software components other than the machine learning model are capable of handling other cross-links, as long as they are not cleavable.

Major Concerns:
Described is only the use of a homo bifunctional crosslinker

We have clarified this in the manuscript.

Minor Concerns:
Is it restricted to Linux systems?

It is not restricted to Linux, but as non-Linux installations are challenging, we solve this issue using docker.

**Reviewer #3:**
Manuscript Summary:
This manuscript introduces the protocol for TX-MS (targeted cross-linking mass spectrometry), a method capable of determining large complex structures, using a combination of high-resolution mass spectrometry, chemical cross-linking, and high-accuracy structure modeling. This method provides an alternative path to solving structures of protein-protein interactions, without requiring time-consuming structural biology experiments such as X-ray crystallography, NMR, and cryo-EM. The protocol introduced here demonstrates the procedures of running TX-MS and interpreting the results, using the

Jupyter Notebook platform.
This manuscript presents its potential importance; however, it lacks sufficient explanations and details for the users to repeat the procedures. In general, the current manuscript might not be qualified for publication in JoVE. After going through the manuscript, I have the following questions that preclude this work from publication in the current form.

We thank the reviewer for the comments and present an improved manuscript which we hope fulfills the requirements for publication in JoVE.

Major Concerns:
1. The title chosen by the authors is neither clear nor informative. The authors should use more general terms for the audiences to comprehend the key purpose of this work from the title. In the same vein, there are plenty of abbreviations running through the article without definition. It would be much more formal and apt to provide the full name of a term before using its abbreviation in the first instance; for example, TX-MS, XL, etc.

We have corrected this. TX-MS is the name of a method in this case.

2. Several sentences used in the abstract and introduction are vague. For instance, in the sentence "removing one of the major bottlenecks," the authors do not clearly explain what these bottlenecks are. In another sentence, "A graphical report summarizing the most important results," what do the authors mean by "the most important results?" More specific explanations would be appreciated.

We have clarified these vague statements.

3. In the section of PROTOCOL, the authors might over-simplify the steps for the first-time user to run TX-MS. First, there is no 'upload data' option on the page so I tried to click 'sign up' option followed by filling up the information to start with the tool. Unfortunately, the server directed to a strange page showing some texts (might be php codes), so I couldn't continue to review this work, even though I tried different web browsers (Chrome and Internet Explore).

We apologize for the inconvenience and have fixed the registration page. We also provided JoVE with login credentials to be used by the reviewers, but something must have gone wrong, and these were not communicated. The login credentials are:

reviewer1 VaayMM1D
reviewer2 XM2CCipW
reviewer3 5rpu4TBi
reviewer4 8UHrPWzq
reviewer5 WG29f6yv

4. If the authors can provide clear information about the version of the tools, or even the operating system used in this demonstration, that would be very useful. After all, there are many cases where the users cannot repeat the experiment due to conflicting versions of the tools or operating system. Also, is there any requirement for the Jupyter Notebook version to be installed before installing the JupyterHub? The authors should clarify all this.

We have clarified the versions of the tools and operating systems.

5. That would be helpful if the authors can note the way how users can run the Jupyter Notebooks on a public IP server, rather then a private computer using http://127.0.0.1 as the default address.

We have clarified this statement.

Minor Concerns:
Several grammatical errors need to be corrected.

We have corrected spelling and grammatical issues.

# Rebuttal letter

Editorial comments

Please take this opportunity to thoroughly proofread the manuscript to ensure that there are no spelling or grammar issues.

We have gone over the manuscript in detail and removed spelling and grammatical issues.

2. Lines 191-231: Please write these as numbered steps.

We added numbered steps to the requested lines.

Reviewer 4

The protocol is already improved compare to the initial submission but in the line 302 the user could mention SIM-XL search engine (http://dx.doi.org/10.1016/j.jprot.2015.01.013), since this the only XL tool that produces a protein-protein interaction map as result, and can help in this protocol to validate the results.

We have added the requested reference.

Reviewer 5

As fairly pointed out by the authors, the major concern of the work is on the limitation to DSS (K-K) cross-linker only. This will restrict greatly the use of the tool presented here by the community as more and more new cross-links are developed and used. This work is another docking tool integrating cross-links data as several other exist and I am not convinced it will be largely used by the community because of its specificity.

We appreciate the reviewer's comments; adding support for other cross-linking reagents is technically possible but early work has indicated that we need to retrain the machine learning model used in the hrMS1 step. We will add support for all cross-linkers that we have sufficient amounts of data to train and validate the models.

There is quite a number of spelling error in the manuscript that need to be corrected. For example :

line 69 "and" should be removed

Removed.

line 294 "cross=linked" instead of "cross-linked",

Corrected.

line 300 "sessility" instead of "sensibility"

Corrected.

Jupyter notebook

Click here to access/download

**Supplemental Coding Files**

Supplementary Data 1.ipynb