Please cite this article in press as: Leitner et al., Toward Increased Reliability, Transparency, and Accessibility in Cross-linking Mass Spectrometry, Structure (2020), https://doi.org/10.1016/j.str.2020.09.011

Structure

CellPress

Meeting Report

Toward Increased Reliability, Transparency, and Accessibility in Cross-linking Mass Spectrometry

Alexander Leitner,[1,*] Alexandre M.J.J. Bonvin,[2] Christoph H. Borchers,[3,4,5,6,7] Robert J. Chalkley,[8] Julia Chamot-Rooke,[9] Colin W. Combe,[10] Jürgen Cox,[11] Meng-Qiu Dong,[12] Lutz Fischer,[13] Michael Götze,[1] Fabio C. Gozzo,[14] Albert J.R. Heck,[15,16] Michael R. Hoopmann,[17] Lan Huang,[18] Yasushi Ishihama,[19] Andrew R. Jones,[20] Nir Kalisman,[21] Oliver Kohlbacher,[22,23,24,25] Karl Mechtler,[26] Robert L. Moritz,[17] Eugen Netz,[22,25] Petr Novak,[27] Evgeniy Petrotchenko,[5,7] Andrej Sali,[28] Richard A. Scheltema,[15,16] Carla Schmidt,[29] David Schriemer,[30,31,32] Andrea Sinz,[33] Frank Sobott,[34,35] Florian Stengel,[36,37] Konstantinos Thalassinos,[38,39] Henning Urlaub,[40] Rosa Viner,[41] Juan A. Vizcaíno,[42] Marc R. Wilkins,[43] and Juri Rappsilber[10,13,*]

[1]Department of Biology, Institute of Molecular Systems Biology, ETH Zurich, 8093 Zurich, Switzerland
[2]Bijvoet Centre for Biomolecular Research, Faculty of Science-Chemistry, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands
[3]University of Victoria-Genome British Columbia Proteomics Centre, University of Victoria, Victoria, BC V8Z 7X8, Canada
[4]Department of Biochemistry and Microbiology, University of Victoria, Victoria, BC V8P 5C2, Canada
[5]Segal Cancer Proteomics Centre, Lady Davis Institute, Jewish General Hospital, McGill University, Montreal, QC H3T 1E2, Canada
[6]Gerald Bronfman Department of Oncology, Jewish General Hospital, McGill University, Montreal, QC H3T 1E2, Canada
[7]Center for Computational and Data-Intensive Science and Engineering, Skolkovo Institute of Science and Technology, Moscow 121205, Russia
[8]Department of Pharmaceutical Chemistry, University of California San Francisco, 600 16th Street, Genentech Hall, Suite N472A, San Francisco, CA 94143-2240, USA
[9]Institut Pasteur-CNRS USR 2000, Mass Spectrometry for Biology Unit, 28, Rue du Dr. Roux, 75 015 Paris, France
[10]Wellcome Centre for Cell Biology, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3BF, Scotland, UK
[11]Computational Systems Biochemistry Research Group, Max-Planck Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany
[12]National Institute of Biological Sciences, Beijing, Beijing 102206, China
[13]Bioanalytics, Institute of Biotechnology, Technische Universität Berlin, 13355 Berlin, Germany
[14]Institute of Chemistry, University of Campinas, Campinas - SP 13083-970, Brazil
[15]Biomolecular Mass Spectrometry and Proteomics, Bijvoet Centre for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, 3584 CH Utrecht, The Netherlands
[16]Netherlands Proteomics Centre, Utrecht University, 3584 CH Utrecht, The Netherlands
[17]Institute for Systems Biology, 401 Terry Avenue North, Seattle, WA 98109, USA
[18]Department of Physiology & Biophysics, University of California, Irvine, Irvine, CA 92697, USA
[19]Graduate School of Pharmaceutical Sciences, Kyoto University, Kyoto 606-8501, Japan
[20]Institute of Integrative Biology, University of Liverpool, Biosciences Building, Crown Street, Liverpool L69 7ZB, UK
[21]Department of Biological Chemistry, Institute of Life Sciences, Hebrew University of Jerusalem, Jerusalem 91904, Israel
[22]Department of Computer Science, University of Tübingen, Sand 14, 72076 Tübingen, Germany
[23]Institute for Bioinformatics and Medical Informatics, University of Tübingen, Sand 14, 72076 Tübingen, Germany
[24]Institute for Translational Bioinformatics, University Hospital Tübingen, Hoppe-Seyler-Strasse 9, 72076 Tübingen, Germany
[25]Biomolecular Interactions, Max Planck Institute for Developmental Biology, Max-Planck-Ring 5, 72076 Tübingen, Germany
[26]Research Institute of Molecular Pathology, Campus-Vienna-Biocenter 1, 1030 Vienna, Austria
[27]Biotechnology and Biomedicine Centre of the Academy of Sciences and Charles University, Institute of Microbiology, The Czech Academy of Sciences, 252 50 Vestec, Czech Republic
[28]Department of Bioengineering and Therapeutic Sciences, Department of Pharmaceutical Chemistry, California Institute for Quantitative Biosciences, University of California, San Francisco, 1700 4th Street, San Francisco, CA 94143, USA
[29]Interdisciplinary Research Center HALOmem, Institute for Biochemistry and Biotechnology, Martin Luther University Halle-Wittenberg, Kurt-Mothes-Strasse 3a, 06120 Halle/Saale, Germany
[30]Department of Biochemistry and Molecular Biology, University of Calgary, Calgary, AB T2N 4N1, Canada
[31]Robson DNA Science Centre, Arnie Charbonneau Cancer Institute, University of Calgary, Calgary, AB T2N 4N1, Canada
[32]Department of Chemistry, University of Calgary, Calgary, AB T2N 4N1, Canada
[33]Department of Pharmaceutical Chemistry & Bioanalytics, Charles Tanford Protein Center, Martin Luther University Halle-Wittenberg, Kurt-Mothes-Strasse 3a, 06120 Halle/Saale, Germany
[34]School of Molecular and Cellular Biology, Faculty of Biological Sciences and Astbury Centre for Structural and Molecular Biology, University of Leeds, Leeds LS2 9JT, UK
[35]Department of Chemistry, Biomolecular and Analytical Mass Spectrometry Group, University of Antwerp, Groenenborgerlaan 171, 2020 Antwerp, Belgium
[36]University of Konstanz, Department of Biology, Universitätsstrasse 10, 78457 Konstanz, Germany
[37]Konstanz Research School Chemical Biology, University of Konstanz, Universitätsstrasse 10, 78457 Konstanz, Germany
[38]Institute of Structural and Molecular Biology, Birkbeck College, University of London, Malet Street, London WC1E 7HX, UK
[39]Institute of Structural and Molecular Biology, Division of Biosciences, University College London, Gower Street, London WC1E 6BT, UK
[40]Bioanalytical Mass Spectrometry, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany
[41]Thermo Fisher Scientific, 355 River Oaks Parkway, San Jose CA 95134, USA
[42]European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK
[43]School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney 2052, Australia
*Correspondence: leitner@imsb.biol.ethz.ch (A.L.), juri.rappsilber@tu-berlin.de (J.R.)
https://doi.org/10.1016/j.str.2020.09.011

Please cite this article in press as: Leitner et al., Toward Increased Reliability, Transparency, and Accessibility in Cross-linking Mass Spectrometry, Structure (2020), https://doi.org/10.1016/j.str.2020.09.011

SUMMARY

Cross-linking mass spectrometry (MS) has substantially matured as a method over the past 2 decades through parallel development in multiple labs, demonstrating its applicability to protein structure determination, conformation analysis, and mapping protein interactions in complex mixtures. Cross-linking MS has become a much-appreciated and routinely applied tool, especially in structural biology. Therefore, it is timely that the community commits to the development of methodological and reporting standards. This white paper builds on an open process comprising a number of events at community conferences since 2015 and identifies aspects of Cross-linking MS for which guidelines should be developed as part of a Cross-linking MS standards initiative.

INTRODUCTION

Cross-linking for structural analysis goes back at least as far as 1958, when the topology of insulin was investigated with the help of a cross-linking reagent (Zahn and Meienhofer, 1958). Introducing mass spectrometry (MS) for the detection of cross-links (cross-linking mass spectrometry, here abbreviated as Cross-linking MS, but also known as XL-MS, CXMS, or CLMS) led to increased accuracy of identifying which pairs of proteins were linked together in heteromeric complexes and increased resolution by revealing the identity of the linked residues and, thus, the interaction regions within these proteins. Technical progress, including the wide variety of parallel developments, and biological applications have been reviewed extensively in recent years (Leitner et al., 2016; O'Reilly and Rappsilber, 2018; Sinz, 2018; Steigenberger et al., 2020; Yu and Huang, 2018).

Encouraged by the Worldwide Protein Data Bank (wwPDB) Task Force for Integrative/Hybrid Methods (Berman et al., 2019; Sali et al., 2015) to provide experimentalists and modelers with a stable access point to cross-linking data, an initial open gathering on standards in the Cross-linking MS field took place at the 14th Human Proteome Organization (HUPO) World Congress 2015 in Vancouver (Canada). This effort was carried forward into an open podium discussion at the 5th Symposium on Structural Proteomics in Halle/Saale (Germany) later that year. At the HUPO-Proteomics Standards Initiative (PSI) meeting 2016 in Ghent (Belgium), an agreement was reached to support cross-linking data starting from v.1.2 of mzIdentML, the proteomics data standard for peptide/protein identification information (Vizcaíno et al., 2017). Following a closed meeting of senior investigators at the 7th Symposium on Structural Proteomics in Vienna (Austria), 2017, again an open podium discussion took place at the 8th Symposium on Structural Proteomics in Berlin (Germany), 2018. A first community-wide, comparative Cross-linking MS study was published in 2019 (Iacobucci et al., 2019), organized through the European Union COST Action BM1403 as an initiative to develop activities in structural proteomics at large, including Cross-linking MS. Discussions were continued during three meetings in 2019: the American Society for Mass Spectrometry Sanibel Conference 2019 entitled "Chemical Cross-linking and Covalent Labeling: From Proteins to Cellular Networks," the Dagstuhl Seminar 19351 "Computational Proteomics," and the 18th HUPO World Congress 2019 in Adelaide. These efforts were brought together at the 9th Symposium on Structural Proteomics in Göttingen (Germany) later that year. A questionnaire on the standardization of Cross-linking MS was circulated among the participants of the meeting. Also, a discussion group consisting of 20 research labs and companies with strong interests in Cross-linking MS, and thus a strong representation of the field, formulated challenges and recommendations for the field. These were publicly discussed within the conference at the end of the meeting. The resulting draft document was then circulated to participants and additional labs in Cross-linking MS that were not represented in Göttingen to provide them with an opportunity to participate. Following the discussions at these meetings, this white paper is now supported by about 30 academic laboratories and companies engaged in developing, applying, and supporting Cross-linking MS, therefore representing a substantial fraction of the field.

CROSS-LINKING AT THE INTERFACE OF PROTEOMICS AND STRUCTURAL BIOLOGY

The definition of Cross-linking MS is as follows: non-covalent interactions or proximities within or between biomolecules are covalently fixed for their detection in an otherwise dissociative analytical process involving a mass spectrometer.

Cross-linking MS shares some similarities with conventional structural biology techniques, but also has some distinct features. For example, structural dynamics in solution is not appropriately reflected in static structures obtained by X-ray crystallography. NMR spectroscopy and, to some extent, (cryo-) electron microscopy (EM) are able to reveal ensembles of conformational states. Cross-linking data also reflect such solution-phase dynamics and are often able to provide crucial contact information about flexible regions in proteins that remain inaccessible to EM or crystallography and are therefore absent in many deposited structures and models. Due to this complementarity to established structural methods, Cross-linking MS has gained acceptance in the structural biology community, and efforts toward standardization should also align with best practices in that field.

Cross-linking MS is most intimately connected to a wide range of applications in structural biology, but is at the same time rooted in MS-based proteomics, two fields where substantial efforts in standardization and harmonization have developed in the past decades (Berman et al., 2006; Burley et al., 2017; Deutsch et al., 2017; Lawson et al., 2011; Montelione et al., 2013; Sali et al., 2015; Schwede et al., 2009; Trewhella et al., 2013, 2017; Vallat et al., 2018). However, the primary data output from Cross-linking MS experiments already differs substantially from conventional protein identification and quantification workflows in proteomics. Instead of identifying single peptide chains that then jointly identify proteins, the main types of identification

Please cite this article in press as: Leitner et al., Toward Increased Reliability, Transparency, and Accessibility in Cross-linking Mass Spectrometry, Structure (2020), https://doi.org/10.1016/j.str.2020.09.011

CellPress

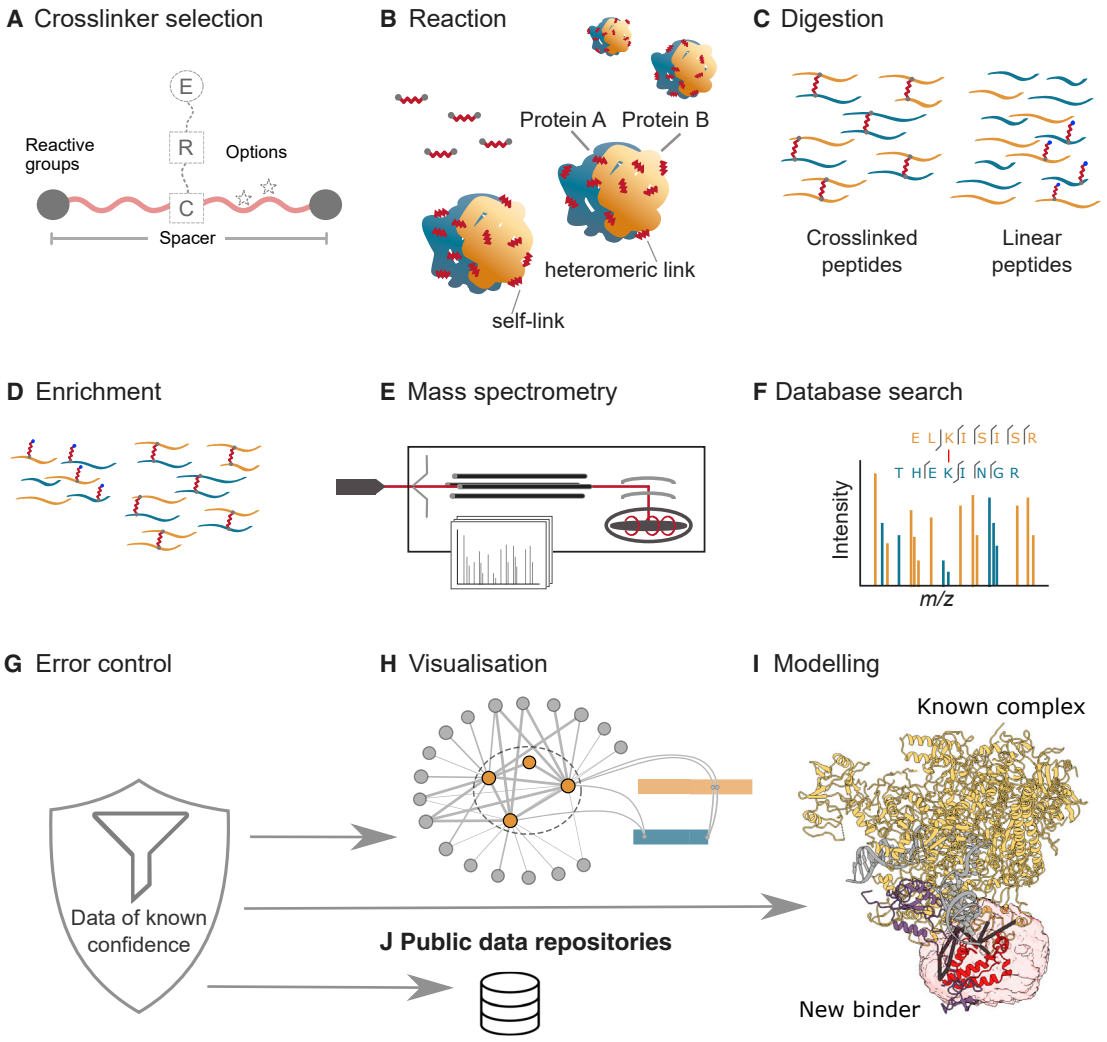

Figure 1. General Cross-linking MS Workflow

(A) Cross-linkers comprise various chemistries and spacer lengths. Depending on the experimental workflow used, the cross-linker spacer may be cleavable in the mass spectrometer (C) or isotope labeled (star) or have moieties that can be biochemically enriched (E) and chemically released (R).

(B) Concentrations and reaction times must be empirically tested for each application to achieve optimal amounts of cross-linking.

(C) Proteins can be digested in solution or in gel to produce a mixture of cross-linked and linear peptides. Also "dead-end" products, where the cross-linker has hydrolyzed on one end; "loop-links," where the cross-link ends up in a single peptide; and higher-order products, comprising more than two peptides and/or more than one cross-linker moiety, can form.

(D) After digestion, cross-linked peptides are often enriched through chromatographic methods, such as size-exclusion chromatography, strong-cation exchange chromatography, or affinity chromatography.

(E) MS/MS acquisition pipelines have been designed to increase the likelihood of selecting cross-linked peptide precursors for fragmentation.

(F) Various search software solutions have been developed to identify the two linked peptides from the spectra.

(G) Through methods that determine the false discovery rate (FDR), the list of matches is cut to the desired confidence.

(H and I) The links are visualized and/or (I) used as part of integrative modeling.

(J) The data are deposited in public repositories. In part adapted from O'Reilly and Rappsilber, 2018.

in Cross-linking MS are pairs of covalently connected, i.e., cross-linked, peptides, which may originate from the same or two different proteins. These are then combined into sets of cross-linking sites (pairs of cross-linked residues) or, for samples of higher complexity, pairs of interacting proteins. Cross-linking involves the use of one or more of a large variety of cross-linking reagents, data acquisition strategies, and data analysis approaches. Inevitably, this diversity will pose particular challenges when it comes to standardizing workflows and data formats and introducing reporting guidelines for the community to adhere to.

Any initiative that attempts to establish standards and guidelines in relation to cross-linking needs to include most of today's diverse cross-linking community. We are in the fortunate position to witness the growth of this community and one might broadly define members of the cross-linking community as (1) researchers or labs that develop cross-linking chemistries, workflows, software, etc., and (2) those that apply cross-linking to address questions in structural biology, molecular biology, systems biology, and so on. In some cases, the focus of research groups in the Cross-linking MS field may, of course, cover

Please cite this article in press as: Leitner et al., Toward Increased Reliability, Transparency, and Accessibility in Cross-linking Mass Spectrometry, Structure (2020), https://doi.org/10.1016/j.str.2020.09.011

both directions. In addition, cross-linking methods are not restricted to the study of protein-protein interactions and protein conformations, but may also include interactions with other classes of biomolecules, including other biopolymers and small molecules. In fact, natural processes can lead to cross-links, and these products can be analyzed by the tools of Cross-linking MS.

To increase transparency and access to results, the cross-linking community has already established some organizational liaisons, including most prominently with proteomics data repositories (ProteomeXchange Consortium [PXC] partners [Deutsch et al., 2020], in particular PRIDE [Perez-Riverol et al., 2019] and jPOST [Moriya et al., 2019]) and repositories for integrative/hybrid structural biology (particularly PDB-Dev, a prototype repository of the wwPDB for integrative structures [Burley et al., 2017; Vallat et al., 2018]). At the moment, these two types of repositories cover different parts of the cross-linking workflow and are not yet interconnected. In the following, we will outline the different steps of this workflow and how researchers from different communities may benefit from increased reliability, transparency, and access.

THE CROSS-LINKING EXPERIMENT: FROM SAMPLE TO SHARED DATA

Figure 1 outlines the different steps of a cross-linking experiment and highlights specific steps of the procedure. For the sake of this discussion, we will assume that, independent of the sample type and the specific chemistry involved, a protein or protein mixture has been cross-linked and subsequently digested into peptides using one or more proteases, and the resulting peptide mixture has been analyzed by liquid chromatography coupled to tandem MS (LC-MS/MS). This experimental workflow can deliver (1) sites of cross-links that may inform modeling of a protein or protein complex structure and (2) information on which proteins were linked and thus interacting in a possibly highly complex biological mixture. The raw/primary data emerging from an LC-MS/MS experiment are considered the starting point of the data analysis pipeline. Primary data that are generated in vendor-specific formats may first be converted into open file formats suitable for database searching (although some programs may be able to work with vendor-specific formats directly). Either way, a peak list is generated that corresponds to the experimentally acquired MS/MS spectra that are searched against a protein sequence database of interest. This database may contain anything from a few protein sequences of interest up to a whole proteome database.

Successful matches to the experimental spectra are designated either peptide-spectrum matches (PSMs) or cross-link-spectrum matches (CSM, XSM). These matches correspond to the (putative) assignment of the sequence of two peptides connected by a cross-link at defined positions within the peptide sequences. This has implications on error handling, as will be discussed below. Therefore, the term cross-link-spectrum match might be more suitable. CSMs/XSMs may subsequently be collapsed into higher-level contact information: peptide pairs, residue pairs, or protein pairs. It should be noted that Cross-linking MS also has to address the protein inference problem of proteomics (Nesvizhskii and Aebersold, 2005; Rappsilber and Mann, 2002) due to the existence of multiple proteins with overlapping sequences. In addition, as multiple copies of the same protein are present in a sample, cross-links of a protein to itself may be intramolecular or intermolecular. Often one cannot distinguish these self-links without dedicated experimental design or additional considerations (Lima et al., 2018; Taverner et al., 2002).

Identifications at all levels are associated with some error rate. This concerns the identity of the peptides (characterized by the false discovery rate, FDR) and the localization of the cross-linking sites (characterized by the false localization rate, FLR). FDRs can, in principle, be determined by so-called target/decoy search strategies whereby the MS/MS spectra are searched against the target sequence database and a database of non-natural decoy sequences that are typically obtained by reversing or shuffling the sequences contained in the target database (Fischer and Rappsilber, 2017, 2018; Maiolica et al., 2007; Walzthoeni et al., 2012). The frequency of matches to the decoy database is then assumed to be equivalent to the frequency of random hits to the target database; this correlation is used to calculate the FDR. There are several caveats to consider for FDR control in cross-linking: first, the mere fact that a combination of two peptides is identified raises the chance for error compared with single peptide chain identifications. It that suffices that if one of the two peptides is false, then the whole assignment will be false (Trnka et al., 2014). Also, false positives increase proportionally when moving from the CSM/XSM to the peptide pair to the protein pair level (Fischer and Rappsilber, 2017). This error propagation is a result of the typically observed redundancy of true positive hits (multiple CSMs/XSMs per peptide pair, multiple peptide pairs per protein pair), while random false positives by definition are less redundant. Therefore, FDRs need to be controlled at multiple levels, in the same way as for conventional proteomics experiments, when moving from CSMs/XSMs to identified peptides and identified proteins. Second, the search spaces of self and heteromeric cross-links are of different sizes. Therefore, the error of self and heteromeric links must be considered separately (Lenz et al., 2020; Walzthoeni et al., 2012). Third, for samples of limited complexity, when using only a small sequence database, and dependent on the cross-linking chemistry and data analysis strategy, there may be an insufficient number of decoy hits for an accurate determination of FDRs. Most FDR strategies try to model the tail of the false positive score distributions, but in a sparse dataset this boundary is strongly affected by the selection of search parameters such as database composition (inclusion/exclusion of contaminant proteins) or the defined specificity of the cross-linking reagent.

Although the field has already seen substantial progress in FDR control, many commonly used software tools do not yet support FDR control at all levels, and there is no consensus for how FDRs should be collected for the various experimental designs currently applied in cross-linking (Beveridge et al., 2020; Keller et al., 2019a; Yugandhar et al., 2020). When dealing with error rates at the cross-linking site level, the FLR (related to the confidence of correctly assigning the cross-linked residues within the two peptide sequences) also needs to be considered. This is an even more challenging problem because the precise localization of the cross-link sites requires the observation of not just any fragments of the peptides but those that allow the

Please cite this article in press as: Leitner et al., Toward Increased Reliability, Transparency, and Accessibility in Cross-linking Mass Spectrometry, Structure (2020), https://doi.org/10.1016/j.str.2020.09.011

exclusion of alternative sites. Often, the linkage site is assigned based on the known or assumed reactivity of the cross-linker, an approach that fails at least when using photo-cross-linking (Schneider et al., 2018). It is worthwhile to note that cross-links reveal proximity, and it remains to be seen what precision level is required by modeling. Here, Cross-linking MS may differ from post-translational modification (PTM) mapping, for example, where the exact PTM site can be critical for mechanistic studies. Additional challenges arise when pairs of the same peptides, but linked at different sites, co-elute chromatographically.

Eventually, the outcomes of cross-linking experiments are made publicly available through different channels, typically journals and data repositories. Research articles provide experimental details and mostly qualitative, but increasingly also quantitative, cross-linking results (Chen and Rappsilber, 2018), in widely varying degrees of detail. Cross-linking identifications are typically reported in a tabular format in the main article or (more commonly) integrated into the online supporting information section as stand-alone tables or formatted in a joint file with other supplementary data, for example, in PDF format. Although this solution may at least fulfill minimum expectations of data transparency, such a deposition complicates data reuse and reanalysis, especially as the formatting lacks a common standard specifying what essential data should be included. Proteomics data repositories already offer some support for cross-linking datasets; for example, a project can be designated as a cross-linking study in the ProteomeXchange partner repository PRIDE, and all data necessary for a "complete" submission can be provided. However, at the time of writing, the submission is labeled as "partial" and therefore may give the false impression of not entirely adhering to open-data-sharing principles and cannot be cited through a digital object identifier, which is increasingly becoming part of open data policies (Gierasch et al., 2020). The PXC partner repository jPOST accepts such submissions as "complete," fortunately. Version 1.2 of the open mzIdentML standard (Vizcaíno et al., 2017), developed by the HUPO-PSI and the official PSI validator, offers support for some cross-linking results (Montecchi-Palazzi et al., 2009), but not all workflows are supported; for example, the increasingly popular cleavable cross-linking reagents are not completely covered. The less complex, tab-delimited mzTab file format would be an alternative.

Apart from the MS-centric data deposition, integrating cross-linking data into other resources, such as protein sequence databases (e.g., UniProt; UniProt Consortium, 2019) or protein interaction databases such as IntAct (Orchard et al., 2014), STRING (Szklarczyk et al., 2019), BioGRID (Oughtred et al., 2019), or Complex Portal (Meldal et al., 2019), would be beneficial. In fact, IntAct is already including published cross-linking data on protein-protein interactions, even though the Cross-linking MS field has not established appropriate quality control mechanisms. Some cross-linking data are also available through individual lab efforts such as XLinkDB (Keller et al., 2019b). In the specific context of the use of cross-linking data for integrative/hybrid modeling, the data dictionary (Vallat et al., 2018) used by PDB-Dev offers support for cross-linking site-centric distance restraints. However, there is no interoperability between these resources that would seamlessly connect all these different repositories and databases. The following section will explain why this would be highly valuable to different audiences.

REQUIREMENTS FOR MAXIMAL IMPACT OF CROSS-LINKING MS

To maximize the use of Cross-linking MS, its data should be made available adhering to the principles of FAIR (findable, accessible, interoperable, reusable) (Wilkinson et al., 2016). In addition, all of the experimental steps should be transparent to others by providing a sufficient amount of information, defined jointly by the community, and by providing this information in a suitable format in articles and in data repositories. This has been done for multiple other proteomics data types (http://www.psidev.info/miape) (Taylor et al., 2007). However, the different communities that stand to benefit from Cross-linking MS data require different types of data and levels of detail for reuse, and this needs to be considered before planning a course of action.

Peers (wet- and dry-lab scientists working in the cross-linking area) may use data to learn about new developments in the field, to assess the validity of published work, and to reanalyze existing datasets, for example, in the context of software development. For these purposes, detailed information and access to many different files are required. This includes raw/primary MS data and peak lists used for the initial search together with the search configuration and database. It also includes "technical" metadata (including details related to the original search such as software [version] and search parameters) and details about the instrumentation for which a first example of a reporting template already exists (Iacobucci et al., 2019). Finally, one also requires identifications at different levels (CSMs/XSMs, peptide pairs, residue/site pairs, protein pairs), including decoys, details about the FDR control (what approach was used and at which levels FDR control was applied, although this should ideally be standardized), and "biological" metadata (related to the nature of the sample and the experimental design, e.g., replicates or perturbations and sample treatment), together with the link to a publication if applicable.

Structural, computational, or systems biologists are more likely not to work with the raw MS data themselves, but they will rather be interested in using the outcomes of cross-linking experiments for modeling protein conformations, protein complexes, or cellular networks. Consequently, these communities may primarily be interested in the identifications at the residue level or protein-protein interaction level with associated measures of confidence (FDR) and, optionally, abundance. The chemistry of the cross-linking reagent should be well defined regarding reactive sites and spacer length to define appropriate boundaries for cross-link restraints. A stable link to the primary data is required, for example, in a proteomics repository, and the data need to be provided in standardized form (also a wwPDB Integrative/Hybrid Methods Task Force recommendation; Berman et al., 2019; Sali et al., 2015). The data should be findable and experimental details documented, i.e., some basic technical and biological metadata need to be associated with the data together with a link to a more detailed description, ideally a publication.

Finally, *molecular and cell biologists* and other researchers interested in protein interactions in general might be interested

Please cite this article in press as: Leitner et al., Toward Increased Reliability, Transparency, and Accessibility in Cross-linking Mass Spectrometry, Structure (2020), https://doi.org/10.1016/j.str.2020.09.011

Table 1. Recommendations (Single-Sentence Summaries of the Field's To-Do List)

No.	Recommendation
1	Define best practices in experimental design for different applications of Cross-linking MS.
2	Find consensus on procedures to reliably assess error rates for all workflow types and at different levels (site pair to protein pair).
3	Ensure support by and complete integration with proteomics data repositories such as those included in ProteomeXchange.
4	Develop consistent terminology and common vocabularies for metadata annotation of Cross-linking MS datasets.
5	Provide enhanced support for data sharing with community-agreed-upon file formats such as mzIdentML or mzTab.
6	Define minimal requirements for reporting Cross-linking MS data in peer-reviewed publications.
7	Facilitate access to modelers by providing results in formats suitable for structure and model repositories, such as PDB/PDB-Dev.
8	Develop parsers for data integration in interaction databases and develop easily accessible visualization tools.
9	Ensure flexibility for new developments in the field; not all steps need to be standardized as workflows evolve.
10	Organize benchmarking studies for objective comparisons of key experimental and computational steps.
11	Establish minimum reporting standards for reporting new or improved reagents and software tools.

in cross-linking data because they represent binary interactions between proteins and/or specific residues in proteins. For these communities, the biggest value will come from access through an intuitive interface to identifications at the residue level or protein-protein interaction level with associated measures of confidence (FDR). This might be best achieved by the integration of such data into resources (databases) that they normally use, such as IntAct, STRING, or UniProt. A useful point of reference would be the HUPO PSI-MI standard, which records molecular interactions without including the supporting MS data. Either these access points may need to expand their data visualization to include topological information or an additional interface may be needed that provides intuitive access also to residue-level information, akin to what is offered by tools such as xVis, xiNET, and xiVIEW (Combe et al., 2015; Graham et al., 2019; Grimm et al., 2015), or in field databases such as ProXL (Riffle et al., 2016).

In summary, different user bases require a different scope and granularity of the information that is obtained from cross-linking experiments. In any case, the ideal scenario would be a transparent and seamless flow of information to and from all resources connected to cross-linking in standardized formats, raising the question of which parts of the workflow can and should be standardized.

RECOMMENDATIONS AS TO WHERE THE CROSS-LINKING MS FIELD REQUIRES STANDARDIZATION

We feel that the Cross-linking MS field will benefit the most from field-developed standards in four specific areas of Cross-linking MS analyses and reporting, leading to the 11 tasks summarized in Table 1 and presented in detail below.

Workflows/Experiments

Recommendation 1: Best Practice in Experimental Design

Although there is a large diversity of analytical tools and concepts being utilized in Cross-linking MS, they are all based on the same principle of preserving structural information by introducing artificial covalent bonds in and between biomolecules that would otherwise be lost during the mass spectrometric analysis. Therefore, guidelines should be developed to ensure that the resulting data can conclusively be interpreted. This should address fundamental aspects of experimental design such as the number and type of replicates and whether different recommendations are required for different types of experiments, e.g., the analysis of highly purified, individual proteins or small protein complexes versus whole-cell analysis and qualitative versus quantitative experiments. This may include control experiments to address oligomerization or sample integrity. An appropriate mechanism has to be set up that allows for finding where best practice guidelines are needed and developing these guidelines while allowing their continuous adaptation as understanding of Cross-linking MS expands.

Recommendation 2: Error Assessment

It is important to determine the error in Cross-linking MS data by a transparent and thoroughly tested method. There are currently a large number of methods for FDR control that are usually based on the target-decoy approach. In addition, comparisons are made to available high-resolution structures, which has its limitations, as these are also experimental data and describe a static representation of a protein or protein complex. It is of the utmost importance that the field arrives at a consensus for procedures that return a reliable error assessment. It is also important that the limits of these procedures be mapped out. It is hoped that future studies will employ this field-agreed-upon error assessment method in its respective current form. This would be helped by swift integration into the main data analysis workflows by their respective developers. Changes in the procedure must be well documented and thoroughly tested before being implemented.

Data Sharing

Recommendation 3: Public Repositories

All Cross-linking MS data should be shared in an open and stable way in a public repository to provide an identifier such as an accession number and a defined data structure to cross-reference cross-linking datasets in other resources. This repository would require standardized metadata (Recommendation 4) and standardized file formats (Recommendation 5). This is already the case for proteomics data and would just require Cross-linking MS specific adaptations of existing proteomics repositories. The PRIDE repository, as one of the PXC members, has committed to being a partner in this endeavor. Basic functionality for "partial" submissions (raw MS data and some metadata) is already available, but should be considered the minimum baseline for data sharing. For a "complete" submission, some

Please cite this article in press as: Leitner et al., Toward Increased Reliability, Transparency, and Accessibility in Cross-linking Mass Spectrometry, Structure (2020), https://doi.org/10.1016/j.str.2020.09.011

extensions need to be made to better address the Cross-linking MS results. Another PXC member, the jPOST repository, has already accepted several Cross-linking MS projects with "complete" submission, but will continually need to work with the field to universally address the diverse data modalities of Cross-linking MS. Ultimately, the criteria for "complete" submission are Cross-linking MS specific and need to be defined by the field to then be implemented through basic, and in the longer term also more elaborate, checks during submission. Ideally, making "complete" submissions for Cross-linking MS data would be independent of criteria applied to the data of other fields.

The Cross-linking MS data that enter public repositories should receive a quality check at all levels, preferably automatically at the point of uploading. This pertains to the elemental integrity of the files and their adherence to the agreed-upon standard formats, which includes semantic validation and readability by parsers that increase the data availability and reach, consistency tests, and other data quality metrics such as a measure of confidence. For this, appropriate software will need to be developed and maintained in a field effort, in collaboration with data repositories. Results of Cross-linking MS that then enter other repositories should do so together with a measure of confidence (see Recommendation 8).

Recommendation 4: Metadata

All information needed to reproduce Cross-linking MS results must be provided in full. Duplications in locations where this takes place should be minimized, though. A minimal set of critical information that is required for a basic understanding of the Cross-linking MS results should be provided as part of data submission. A standardized description of a cross-linking experiment requires the definition of common, controlled vocabularies. XLMOD (Mayer, 2020) (https://raw.githubusercontent.com/HUPO-PSI/mzIdentML/master/cv/XLMOD.obo) is an effort coordinated via the HUPO-PSI on controlled vocabularies in cross-linking that covers "cross-linking reagents, cross-linker related post-translational modifications, and derivatisation reagents for GC-MS and LC-MS." Other terms for a standardized, minimal description of a Cross-linking MS experiment will need to be defined in additional efforts. Full experimental details should be provided in the experimental section of publications. A good starting point of what should be included here are the recommendations emerging from a first community study (Iacobucci et al., 2019). This is currently taking a tabular form as is also practiced in other fields (Henderson et al., 2012; Masson et al., 2019; Montelione et al., 2013; Read et al., 2011; Trewhella et al., 2017). Metadata about instrumentation and data acquisition parameters would ideally be parsed from the raw data before submission and written automatically into the submitted file. Likewise, search parameters should automatically be documented in a form so that users can effortlessly pass them on to the data repository as part of their submission. It would be desirable to minimize the number of separate files by combining all relevant experimental information together with the results into a single file.

Recommendation 5: Community-Agreed-upon File Formats

All data should be shared in an open and community-agreed-upon format that is extensible to support the evolving needs of the community. This has been successfully performed for peak list formats such as mzML. A standard result file format should be developed and include a complete list of target and decoy identifications as potential true and known noise distributions. Integrative/hybrid modeling is tolerant to substantial error rates, and including the known noise (decoy matches) allows the modeling field to build ways to deal with noise in cross-linking data into their procedures (Berman et al., 2019; Rout and Sali, 2019). From a modeling perspective, such decoy matches represent an initial noise distribution that can be converted into a more accurate noise model by FDR estimation procedures. This format might be based on already existing standard formats such as mzIdentML or mzTab, which are supported by the HUPO-PSI as the initiative in the proteomics field on procedures for standardization. mzIdentML 1.2 would be a starting point for further efforts since it already supports some, albeit not all, types of cross-linking data (Vizcaíno et al., 2017). Cross-linking data are currently not supported in mzTab, but mzTab could be extended to accommodate these data. In addition to output formats, one should also keep an eye on input files and their standardization, which includes mzML for peak lists (Martens et al., 2011) and PEFF for sequence databases (Binz et al., 2019). Part of integrating the needs of the cross-linking community into general proteomics standards will be developing parser libraries, readers, and writers. Cross-linking MS search software should be adapted to write results and search parameters in a standards compliant form to allow direct sharing of data, metadata, and results. As cross-linking is evolving as a method, this will also lead to evolving standards and a continuous need to update software tools. We acknowledge that the complexity of the data makes the whole process of changing existing software tools and maintaining them challenging.

Recommendation 6: Publication Guidelines

Publication guidelines should be developed for what constitutes a sufficiently detailed description of experimental design (Recommendation 1), sample and data processing (Recommendations 2 and 4), and presentation of results (Recommendations 3 and 5).

Knowledge Transfer

Recommendation 7: Access to Cross-linking MS Results for Modeling

Efforts should be undertaken to maximize access of other researchers and communities to the link and interaction data obtained by Cross-linking MS. In fact, wwPDB/PDB-Dev has reached out in the name of the structural biology and modeling field with the specific requirement of having access to Cross-linking MS data held in a public repository in standardized form and with quality descriptors. These requirements are going to be met by Recommendations 2 (FDR), 3 (data repository), and 5 (file formats). Once Cross-linking MS formats have been established, parsers can and will be written to stably link the data into the workflows of the modeling community and the prototype archiving system for structural models, PDB-Dev.

Recommendation 8: Accessibility of Cross-linking MS Data to Experimentalists

Providing biological researchers with efficient access to Cross-linking MS results requires those data to be integrated into existing resources containing protein-protein interaction information, such as IntAct, UniProt, and STRING. This requires, first of all, Cross-linking MS data to make it into these repositories in an

Please cite this article in press as: Leitner et al., Toward Increased Reliability, Transparency, and Accessibility in Cross-linking Mass Spectrometry, Structure (2020), https://doi.org/10.1016/j.str.2020.09.011

automated and quality-controlled way (linking to Recommendation 2 and 3). This includes the writing of parsers that convert Cross-linking MS results into formats for molecular interactions and protein complexes. This also mandates the further development of cross-linking data visualization tools and their integration with public databases. These visualization tools can be broadly categorized by their purpose: investigating spectral data, protein structure, or protein interaction networks. This does not necessarily comprise a definitive listing, as these tools and their integration with one another and other tools are under active development. Even within the first category, consisting of spectral interpretations, the often long lists of cross-linked proteins and cross-linked amino acid residues returned by Cross-linking MS are not intuitively understandable. The ability to display residue-resolution information provided by Cross-linking MS has been shown to provide a more suitable visual data interaction platform (Combe et al., 2015; Graham et al., 2019; Hoopmann et al., 2015, 2016; Keller et al., 2019b; Kolbowski et al., 2018; Riffle et al., 2016, among a subset of examples). Tools increase in value through integration; for example, node-link diagrams that classically display protein interaction data (Combe et al., 2015) can be supplemented by a display of the residue-level spectral interpretations (Graham et al., 2019; Riffle et al., 2016). Visualization tools should be further developed to allow a seamless interrogation of Cross-linking MS data from a wide angle of perspectives, with a focus on understanding the data and developing testable hypotheses. This requires linking the visualization to the public repository of Cross-linking MS data on one side and public repositories of protein function and interaction data on the other side. Visualization tools should have low entry barriers, such as being browser based (Combe et al., 2015; Deutsch et al., 2015; Graham et al., 2019; Kolbowski et al., 2018; Riffle et al., 2016; Trnka et al., 2014) or easy to install (Kosinski et al., 2015), and be open source and grant funded to ensure transparent development and access by the widest possible number of researchers.

Future Development of Cross-linking MS

Recommendation 9: Cross-linking MS Comes in Many Flavors

Cross-linking MS is currently seeing the rapid prototyping of novel workflows. These workflows implement different ideas around the same basic concept but use in part very different analytical tools. Given the diversity of approaches that exist for proteomics, it is unclear if a unified workflow will arise for Cross-linking MS (Leitner et al., 2016; O'Reilly and Rappsilber, 2018; Sinz, 2018; Steigenberger et al., 2020; Yu and Huang, 2018). Therefore, many fundamental elements of the workflow *should not* be subject to strict standardization at this point. This specifically includes (1) cross-linking reagents, as many types of cross-linking reagents and chemistries exist and new ones are introduced on a regular basis; (2) instrumentation, as many types of mass spectrometers with diverse features (e.g., fragmentation techniques, real-time decision-making processes) exist and certainly the technology will continue to evolve; and (3) data analysis software, as many types of cross-linking analysis software exist and, again, there is a continual development of entirely new software tools or new versions of existing tools.

In addition, there are many different ways of combining chemistry, MS, and bioinformatics, although dependencies can and do exist (e.g., a software will work only with cross-linking reagents of a certain design, or it may accept only certain types of MS data). Although some workflows may be more suitable than others for a given application, cross-linking can be applied in many different contexts. There are a large number of strategies being explored and that will continue being explored for the foreseeable future. Note that diversity also exists in other fields; for example, many different software tools are used successfully for protein identification in proteomics.

Recommendation 10: Community Benchmark Exercise

The first community-wide, comparative Cross-linking MS study published in 2019 (Iacobucci et al., 2019) highlighted the desire of the field for transparent assessments of the many different workflows through organized challenges. This should be continued and expanded to include all application areas of Cross-linking MS, from single proteins to multi-protein complexes and to highly complex mixtures of proteins. These challenges should provide both experimentalists and computational scientists the opportunity to showcase and benchmark their tools and demonstrate the progress of the field and/or highlight remaining challenges in the respective areas.

Recommendation 11: Minimal Standards for the Reporting of New Tools

New cross-linking reagents and new search software (versions) are frequently reported. Although these reports typically contain proof-of-principle evidence, they often lack data that allow assessing in full the merits of these new or changed tools. This not only limits the uptake of these tools but also makes it more difficult for others to plan experiments in light of the many choices that are available. We should therefore develop guidelines and possibly benchmark challenges that provide the field with a general comparability of tools and ideally a quantitative assessment of progress. The above-mentioned organized challenges are one approach to this, although infrequent and should be supplemented by rolling and/or fixed challenges. Rolling challenges are known, for example, in protein structure modeling: CAMEO (continuous automated model evaluation) (Haas et al., 2018). Here an available yet confidential structure is used as a ground truth against which submitted models are assessed. Especially when it comes to protein-protein interactions, such a ground truth typically does not exist for Cross-linking MS, and therefore alternative approaches for evaluation will need to be developed.

IMPLEMENTATION

Crowd sourcing community standards is known to be a very time-consuming process. To streamline this process, we suggest that initially a small group gathered from the authors of this paper and any other interested party (please contact A.L. or J.R.) proposes such standards. Initial discussions may be performed through online discussion groups, where different opinions on best practices can be shared and specific challenges discussed. Once a consensus emerges, recommendations would be presented at a community meeting, such as the annual Symposium on Structural Proteomics, and reported in a publication. This process is to some extent reminiscent of the

procedures established by the HUPO-PSI. Many of the different aspects of standardization will require significant funding, for example, to develop new and adapt existing software, and continuous funding support to ensure that standards and their implementation tools evolve with the changing needs of this rapidly developing field. This white paper establishes the foundation of future activities by defining common goals of the Cross-linking MS community. This is an essential first step to show funding bodies that this is not an isolated effort of a single or a few research groups.

ACKNOWLEDGMENTS

We thank the participants of the meetings mentioned throughout the text for their many different contributions. Although we made an effort to include all labs that may feel a belonging to the field of Cross-linking MS, we apologize to all those that we missed. Developing standards in Cross-linking MS is to be viewed as an inclusive activity, so please feel warmly invited to join our efforts. This work was supported by the Wellcome Trust through a Senior Research Fellowship to J.R. (103139) and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, project 329673113). Funded by the DFG under Germany's Excellence Strategy, EXC 2008/1-390540038. The Wellcome Centre for Cell Biology is supported by core funding from the Wellcome Trust (203149).

AUTHOR CONTRIBUTIONS

Writing – Original Draft, A.L. and J.R.; Writing – Review & Editing, all authors.

REFERENCES

Berman, H.M., Burley, S.K., Chiu, W., Sali, A., Adzhubei, A., Bourne, P.E., Bryant, S.H., Dunbrack, R.L., Jr., Fidelis, K., Frank, J., et al. (2006). Outcome of a workshop on archiving structural models of biological macromolecules. Structure *14*, 1211–1217.

Berman, H.M., Adams, P.D., Bonvin, A.A., Burley, S.K., Carragher, B., Chiu, W., DiMaio, F., Ferrin, T.E., Gabanyi, M.J., Goddard, T.D., et al. (2019). Federating structural models and data: outcomes from a workshop on archiving integrative structures. Structure *27*, 1745–1759.

Beveridge, R., Stadlmann, J., Penninger, J.M., and Mechtler, K. (2020). A synthetic peptide library for benchmarking crosslinking-mass spectrometry search engines for proteins and protein complexes. Nat. Commun. *11*, 742.

Binz, P.-A., Shofstahl, J., Vizcaíno, J.A., Barsnes, H., Chalkley, R.J., Menschaert, G., Alpi, E., Clauser, K., Eng, J.K., Lane, L., et al. (2019). Proteomics standards initiative extended FASTA format. J. Proteome Res. *18*, 2686–2692.

Burley, S.K., Kurisu, G., Markley, J.L., Nakamura, H., Velankar, S., Berman, H.M., Sali, A., Schwede, T., and Trewhella, J. (2017). PDB-Dev: a prototype system for depositing integrative/hybrid structural models. Structure *25*, 1317–1318.

Chen, Z.A., and Rappsilber, J. (2018). Protein dynamics in solution by quantitative crosslinking/mass spectrometry. Trends Biochem. Sci. *43*, 908–920.

Combe, C.W., Fischer, L., and Rappsilber, J. (2015). xiNET: cross-link network maps with residue resolution. Mol. Cell. Proteomics *14*, 1137–1147.

Deutsch, E.W., Mendoza, L., Shteynberg, D., Slagel, J., Sun, Z., and Moritz, R.L. (2015). Trans-Proteomic Pipeline, a standardized data processing pipeline for large-scale reproducible proteomics informatics. Proteomics Clin. Appl. *9*, 745–754.

Deutsch, E.W., Orchard, S., Binz, P.-A., Bittremieux, W., Eisenacher, M., Hermjakob, H., Kawano, S., Lam, H., Mayer, G., Menschaert, G., et al. (2017). Proteomics standards initiative: fifteen years of progress and future work. J. Proteome Res. *16*, 4288–4298.

Deutsch, E.W., Bandeira, N., Sharma, V., Perez-Riverol, Y., Carver, J.J., Kundu, D.J., García-Seisdedos, D., Jarnuczak, A.F., Hewapathirana, S., Pullman, B.S., et al. (2020). The ProteomeXchange consortium in 2020: enabling "big data" approaches in proteomics. Nucleic Acids Res. *48*, D1145–D1152.

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

Fischer, L., and Rappsilber, J. (2018). False discovery rate estimation and heterobifunctional cross-linkers. PLoS One *13*, e0196672.

Gierasch, L.M., Davidson, N.O., Rye, K.-A., and Burlingame, A.L. (2020). The data must be accessible to all. Mol. Cell. Proteomics *19*, 569–570.

Graham, M.J., Combe, C., Kolbowski, L., and Rappsilber, J. (2019). xiView: a common platform for the downstream analysis of Crosslinking Mass Spectrometry data. bioRxiv. https://doi.org/10.1101/561829.

Grimm, M., Zimniak, T., Kahraman, A., and Herzog, F. (2015). xVis: a web server for the schematic visualization and interpretation of crosslink-derived spatial restraints. Nucleic Acids Res. *43*, W362–W369.

Haas, J., Barbato, A., Behringer, D., Studer, G., Roth, S., Bertoni, M., Mostaguir, K., Gumienny, R., and Schwede, T. (2018). Continuous Automated Model EvaluatiOn (CAMEO) complementing the critical assessment of structure prediction in CASP12. Proteins *86* (*Suppl 1*), 387–398.

Henderson, R., Sali, A., Baker, M.L., Carragher, B., Devkota, B., Downing, K.H., Egelman, E.H., Feng, Z., Frank, J., Grigorieff, N., et al. (2012). Outcome of the first electron microscopy validation task force meeting. Structure *20*, 205–214.

Hoopmann, M.R., Zelter, A., Johnson, R.S., Riffle, M., MacCoss, M.J., Davis, T.N., and Moritz, R.L. (2015). Kojak: efficient analysis of chemically cross-linked protein complexes. J. Proteome Res. *14*, 2190–2198.

Hoopmann, M.R., Mendoza, L., Deutsch, E.W., Shteynberg, D., and Moritz, R.L. (2016). An open data format for visualization and analysis of cross-linked mass spectrometry results. J. Am. Soc. Mass Spectrom. *27*, 1728–1734.

Iacobucci, C., Piotrowski, C., Aebersold, R., Amaral, B.C., Andrews, P., Bernfur, K., Borchers, C., Brodie, N.I., Bruce, J.E., Cao, Y., et al. (2019). First community-wide, comparative cross-linking mass spectrometry study. Anal. Chem. *91*, 6953–6961.

Keller, A., Chavez, J.D., Felt, K.C., and Bruce, J.E. (2019a). Prediction of an upper limit for the fraction of interprotein cross-links in large-scale in vivo cross-linking studies. J. Proteome Res. *18*, 3077–3085.

Keller, A., Chavez, J.D., Eng, J.K., Thornton, Z., and Bruce, J.E. (2019b). Tools for 3D interactome visualization. J. Proteome Res. *18*, 753–758.

Kolbowski, L., Combe, C., and Rappsilber, J. (2018). xiSPEC: web-based visualization, analysis and sharing of proteomics data. Nucleic Acids Res. *46*, W473–W478.

Kosinski, J., von Appen, A., Ori, A., Karius, K., Müller, C.W., and Beck, M. (2015). Xlink Analyzer: software for analysis and visualization of cross-linking data in the context of three-dimensional structures. J. Struct. Biol. *189*, 177–183.

Lawson, C.L., Baker, M.L., Best, C., Bi, C., Dougherty, M., Feng, P., van Ginkel, G., Devkota, B., Lagerstedt, I., Ludtke, S.J., et al. (2011). EMDataBank.org: unified data resource for CryoEM. Nucleic Acids Res. *39*, D456–D464.

Leitner, A., Faini, M., Stengel, F., and Aebersold, R. (2016). Crosslinking and mass spectrometry: an integrated technology to understand the structure and function of molecular machines. Trends Biochem. Sci. *41*, 20–32.

Lenz, S., Sinn, L.R., O'Reilly, F.J., Fischer, L., Wegner, F., and Rappsilber, J. (2020). Reliable identification of protein-protein interactions by crosslinking mass spectrometry. bioRxiv. https://doi.org/10.1101/2020.05.25.114256.

Lima, D.B., Melchior, J.T., Morris, J., Barbosa, V.C., Chamot-Rooke, J., Fioramonte, M., Souza, T.A.C.B., Fischer, J.S.G., Gozzo, F.C., Carvalho, P.C., et al. (2018). Characterization of homodimer interfaces with cross-linking mass spectrometry and isotopically labeled proteins. Nat. Protoc. *13*, 431–458.

Maiolica, A., Cittaro, D., Borsotti, D., Sennels, L., Ciferri, C., Tarricone, C., Musacchio, A., and Rappsilber, J. (2007). Structural analysis of multiprotein complexes by cross-linking, mass spectrometry, and database searching. Mol. Cell. Proteomics *6*, 2200–2211.

Martens, L., Chambers, M., Sturm, M., Kessner, D., Levander, F., Shofstahl, J., Tang, W.H., Römpp, A., Neumann, S., Pizarro, A.D., et al. (2011). mzML—a community standard for mass spectrometry data. Mol. Cell. Proteomics *10*, R110.000133.

Masson, G.R., Burke, J.E., Ahn, N.G., Anand, G.S., Borchers, C., Brier, S., Bou-Assaf, G.M., Engen, J.R., Englander, S.W., Faber, J., et al. (2019). Recommendations for performing, interpreting and reporting hydrogen deuterium exchange mass spectrometry (HDX-MS) experiments. Nat. Methods *16*, 595–602.

Mayer, G. (2020). XLMOD: cross-linking and chromatography derivatization reagents ontology. arXiv https://arxiv.org/abs/2003.00329.

Meldal, B.H.M., Bye-A-Jee, H., Gajdoš, L., Hammerová, Z., Horácková, A., Melicher, F., Perfetto, L., Pokorný, D., Lopez, M.R., Türková, A., et al. (2019). Complex Portal 2018: extended content and enhanced visualization tools for macromolecular complexes. Nucleic Acids Res. *47*, D550–D558.

Montecchi-Palazzi, L., Kerrien, S., Reisinger, F., Aranda, B., Jones, A.R., Martens, L., and Hermjakob, H. (2009). The PSI semantic validator: a framework to check MIAPE compliance of proteomics data. Proteomics *9*, 5112–5119.

Montelione, G.T., Nilges, M., Bax, A., Güntert, P., Herrmann, T., Richardson, J.S., Schwieters, C.D., Vranken, W.F., Vuister, G.W., Wishart, D.S., et al. (2013). Recommendations of the wwPDB NMR validation task force. Structure *21*, 1563–1570.

Moriya, Y., Kawano, S., Okuda, S., Watanabe, Y., Matsumoto, M., Takami, T., Kobayashi, D., Yamanouchi, Y., Araki, N., Yoshizawa, A.C., et al. (2019). The jPOST environment: an integrated proteomics data repository and database. Nucleic Acids Res. *47*, D1218–D1224.

Nesvizhskii, A.I., and Aebersold, R. (2005). Interpretation of shotgun proteomic data: the protein inference problem. Mol. Cell. Proteomics *4*, 1419–1440.

Orchard, S., Ammari, M., Aranda, B., Breuza, L., Briganti, L., Broackes-Carter, F., Campbell, N.H., Chavali, G., Chen, C., del-Toro, N., et al. (2014). The MIntAct project–IntAct as a common curation platform for 11 molecular interaction databases. Nucleic Acids Res. *42*, D358–D363.

O'Reilly, F.J., and Rappsilber, J. (2018). Cross-linking mass spectrometry: methods and applications in structural, molecular and systems biology. Nat. Struct. Mol. Biol. *25*, 1000–1008.

Oughtred, R., Stark, C., Breitkreutz, B.-J., Rust, J., Boucher, L., Chang, C., Kolas, N., O'Donnell, L., Leung, G., McAdam, R., et al. (2019). The BioGRID interaction database: 2019 update. Nucleic Acids Res. *47*, D529–D541.

Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D.J., Inuganti, A., Griss, J., Mayer, G., Eisenacher, M., et al. (2019). The PRIDE database and related tools and resources in 2019: improving support for quantification data. Nucleic Acids Res. *47*, D442–D450.

Rappsilber, J., and Mann, M. (2002). What does it mean to identify a protein in proteomics? Trends Biochem. Sci. *27*, 74–78.

Read, R.J., Adams, P.D., Arendall, W.B., 3rd, Brunger, A.T., Emsley, P., Joosten, R.P., Kleywegt, G.J., Krissinel, E.B., Lütteke, T., Otwinowski, Z., et al. (2011). A new generation of crystallographic validation tools for the protein data bank. Structure *19*, 1395–1412.

Riffle, M., Jaschob, D., Zelter, A., and Davis, T.N. (2016). 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.

Rout, M.P., and Sali, A. (2019). Principles for integrative structural biology studies. Cell *177*, 1384–1403.

Sali, A., Berman, H.M., Schwede, T., Trewhella, J., Kleywegt, G., Burley, S.K., Markley, J., Nakamura, H., Adams, P., Bonvin, A.M.J.J., et al. (2015). Outcome of the first wwPDB hybrid/integrative methods task force workshop. Structure *23*, 1156–1167.

Schneider, M., Belsom, A., and Rappsilber, J. (2018). Protein tertiary structure by crosslinking/mass spectrometry. Trends Biochem. Sci. *43*, 157–169.

Schwede, T., Sali, A., Honig, B., Levitt, M., Berman, H.M., Jones, D., Brenner, S.E., Burley, S.K., Das, R., Dokholyan, N.V., et al. (2009). Outcome of a workshop on applications of protein models in biomedical research. Structure *17*, 151–159.

Sinz, A. (2018). Cross-linking/mass spectrometry for studying protein structures and protein-protein interactions: where are we now and where should we go from here? Angew. Chem. Int. Ed. Engl. *57*, 6390–6396.

Steigenberger, B., Albanese, P., Heck, A.J.R., and Scheltema, R.A. (2020). To cleave or not to cleave in XL-MS? J. Am. Soc. Mass Spectrom. *31*, 196–206.

Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P., et al. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. *47*, D607–D613.

Taverner, T., Hall, N.E., O'Hair, R.A.J., and Simpson, R.J. (2002). Characterization of an antagonist interleukin-6 dimer by stable isotope labeling, cross-linking, and mass spectrometry. J. Biol. Chem. *277*, 46487–46492.

Taylor, C.F., Paton, N.W., Lilley, K.S., Binz, P.-A., Julian, R.K., Jr., Jones, A.R., Zhu, W., Apweiler, R., Aebersold, R., Deutsch, E.W., et al. (2007). The minimum information about a proteomics experiment (MIAPE). Nat. Biotechnol. *25*, 887–893.

Trewhella, J., Hendrickson, W.A., Kleywegt, G.J., Sali, A., Sato, M., Schwede, T., Svergun, D.I., Tainer, J.A., Westbrook, J., and Berman, H.M. (2013). Report of the wwPDB Small-Angle Scattering Task Force: data requirements for biomolecular modeling and the PDB. Structure *21*, 875–881.

Trewhella, J., Duff, A.P., Durand, D., Gabel, F., Guss, J.M., Hendrickson, W.A., Hura, G.L., Jacques, D.A., Kirby, N.M., Kwan, A.H., et al. (2017). 2017 publication guidelines for structural modelling of small-angle scattering data from biomolecules in solution: an update. Acta Crystallogr. D Struct. Biol. *73*, 710–728.

Trnka, M.J., Baker, P.R., Robinson, P.J.J., Burlingame, A.L., and Chalkley, R.J. (2014). Matching cross-linked peptide spectra: only as good as the worse identification. Mol. Cell. Proteomics *13*, 420–434.

UniProt Consortium (2019). UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res. *47*, D506–D515.

Vallat, B., Webb, B., Westbrook, J.D., Sali, A., and Berman, H.M. (2018). Development of a prototype system for archiving integrative/hybrid structure models of biological macromolecules. Structure *26*, 894–904.e2.

Vizcaíno, J.A., Mayer, G., Perkins, S., Barsnes, H., Vaudel, M., Perez-Riverol, Y., Ternent, T., Uszkoreit, J., Eisenacher, M., Fischer, L., et al. (2017). The mzIdentML data standard version 1.2, supporting advances in proteome informatics. Mol. Cell. Proteomics *16*, 1275–1285.

Walzthoeni, T., Claassen, M., Leitner, A., Herzog, F., Bohn, S., Förster, F., Beck, M., and Aebersold, R. (2012). False discovery rate estimation for cross-linked peptides identified by mass spectrometry. Nat. Methods *9*, 901–903.

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data *3*, 160018.

Yu, C., and Huang, L. (2018). Cross-linking mass spectrometry: an emerging technology for interactomics and structural biology. Anal. Chem. *90*, 144–165.

Yugandhar, K., Wang, T.-Y., Leung, A.K.-Y., Lanz, M.C., Motorykin, I., Liang, J., Shayhidin, E.E., Smolka, M.B., Zhang, S., and Yu, H. (2020). MaXLinker: proteome-wide cross-link identifications with high specificity and sensitivity. Mol. Cell. Proteomics *19*, 554–568.

Zahn, V.H., and Meienhofer, J. (1958). Reaktionen von 1,5-difluor-2,4-dinitrobenzol mit insulin 2. Mitt. Versuche mit insulin. Makromol. Chem. *26*, 153–166.