Received: 14 July 2017 | Revised: 12 October 2017 | Accepted: 9 January 2018
DOI: 10.1002/mas.21559

REVIEW ARTICLE

WILEY MassSpectrometryReviews

Cross-linked peptide identification: A computational forest of algorithms

Şule Yılmaz[1,2] | Genet A. Shiferaw[1,2] | Josep Rayo[3] | Anastassios Economou[3] | Lennart Martens[1,2] | Elien Vandermarliere[1,2]

[1] VIB-UGent Center for Medical Biotechnology, VIB, Ghent, Belgium

[2] Department of Biochemistry, Ghent University, Ghent, Belgium

[3] Department of Microbiology and Immunology, Rega Institute for Medical Research, Laboratory of Molecular Bacteriology, KULeuven—University of Leuven, Leuven, Belgium

Correspondence
Lennart Martens, VIB-UGent Center for Medical Biotechnology, A. Baertsoenkaai 3, Gent B-9000, Belgium.
Email: lennart.martens@vib-ugent.be

Funding information
RiMembR Vlaanderen Onderzoeksprojecten, Grant numbers: #G0C6814N, FWO; T3RecS, Grant numbers: #G002516N, FWO; DIP-BiD, Grant numbers: #AKUL/15/40 – G0H2116N, Hercules/FWO; R&D Pilot project-2015; Concerted Research Action from Ghent University, Grant number: BOF12/GOA/014

Chemical cross-linking analyzed by mass spectrometry (XL-MS) has become an important tool in unravelling protein structure, dynamics, and complex formation. Because the analysis of cross-linked proteins with mass spectrometry results in specific computational challenges, many computational tools have been developed to identify cross-linked peptides from mass spectra and subsequently interpret the identified cross-links within their structural context. In this review, we will provide an overview of the different tools that are currently available to tackle the computational part of an XL-MS experiment. First, we give an introduction on the computational challenges encountered when processing data from a cross-linking experiment. We then discuss available tools to identify peptides that are linked by intact or MS-cleavable cross-linkers, and we provide an overview of tools to interpret cross-linked peptides in the context of protein structure. Finally, we give an outlook on data management and dissemination challenges and opportunities for cross-linking experiments.

KEYWORDS
algorithms, cross-linking, identification, mass spectrometry, proteomics

1 | INTRODUCTION

Detailed knowledge about the structure of a protein is an important component in resolving its properties and function. One of the main approaches to obtain such structural knowledge of a protein is X-ray crystallography. The result, however, is a static image of a protein while proteins are dynamic entities that can undergo—sometimes dramatic—conformational changes that are essential for their biological function and for their interaction with multiple other protein partners.[1] X-ray crystallography is hence not able to capture the full conformational repertoire of a protein. Both nuclear magnetic resonance (NMR) and hydrogen-deuterium exchange (HDX) provide information on flexibility but have their own limitations.[2,3] Nowadays, when researchers wish to obtain as much structural information as possible, they therefore typically resort to integrative structural biology, which combines several approaches.[4] One frequently used approach in such integrative structural biology investigations is chemical cross-linking analyzed by mass spectrometry (XL-MS). Results from XL-MS analyses can be combined with results from X-ray crystallography, NMR, or cryo-electron microscopy and can thereby bridge for example the gap between low resolution electron microscopy images and the structures of single subunits as determined by high resolution methods.[5–7]

Chemical cross-linking experiments were already performed in the 1970s.[8] At that time, they were used in conjunction with X-ray crystallography. Identification of cross-linked peptides and more specifically, of the cross-linked residues, via mass spectrometry was very difficult at the time due to the absence of soft ionization methods. In the last few years, however, the field of XL-MS has progressed rapidly thanks to improvements in mass spectrometers, cross-linking experimental designs, and bioinformatics tools.[9]

Currently address: Şule Yılmaz, Computational Systems Biochemistry, Max-Planck Institute for Biochemistry, D-82152 Martinsried, Germany.

 © 2018 Wiley Periodicals, Inc.

The experimental setup of a cross-linking experiment is reasonably straightforward; it is the subsequent bio-informatics processing of the data acquired in the experiment that forms the bottleneck for cross-linking studies.[10] Several tools have therefore been developed for the identification of cross-linked peptides; many groups in the field have built their own algorithms that are often tailored to their specific cross-linking strategy, or even to the specific cross-linker that is used within that group.[11]

In this review, we provide an overview of the currently available algorithms and bioinformatics approaches to tackle the computational part of an XL-MS experiment.

2 | BRIEF OVERVIEW OF THE CHALLENGES IN CROSS-LINKED PEPTIDE ANALYSIS

The experimental setup of an XL-MS experiment and the subsequent measurement of the generated peptides are very similar to a regular shotgun proteomics experiment (Figure 1). It is the subsequent identification of the cross-linked peptides that is much more challenging than the identification of single peptides.

The first challenge is the increased complexity of the peptide mixture. The mixture can contain up to five different types of peptides.[9,12] Figure 2 illustrates the different types of peptides and their nomenclature as proposed by Schilling[12] and extended by Rappsilber.[9] The simplest form of peptides that is present in the mixture consists of ordinary or unmodified linear peptides that are not attached to a cross-linker; these are essentially identical to the peptides found in normal shotgun proteomics experiments. Type 0 peptides, frequently referred to as dead-end peptides, are peptides that are attached to one arm of the cross-linker, while the other arm of the cross-linker remains unattached. This type of linkage does not provide distance information but it does provide information on the solvent accessibility of the modified residue.[13] Type 1 peptides, or loop-linked peptides, contain an intra peptide cross-link. These peptides have two of their residues attached to either end of the same cross-linker. Type 2 peptides are two peptides that are linked by a single cross-linker. These peptides can either originate from the same protein (intra protein cross-link) or from different proteins (inter protein cross-link). Intra protein links provide information about the position of domains within a protein[14] and can be used to guide or validate homology modeling studies.[15] Inter protein links are important in the delineation of protein contact interfaces and they provide valuable information about the position of the subunits within a complex.[16] Type 0, type 1, and type 2 peptides are all based on a single modification. Type 3 peptides or higher order peptides are a combination of the other peptide types and hence contain several linkages.[9]

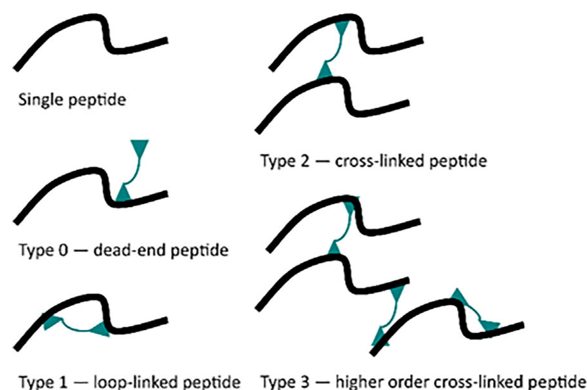

FIGURE 2 Nomenclature of the different types of peptides found in a cross-linking experiment. This nomenclature is based on the nomenclature proposed by Schilling[12] and extended by Rappsilber[9]

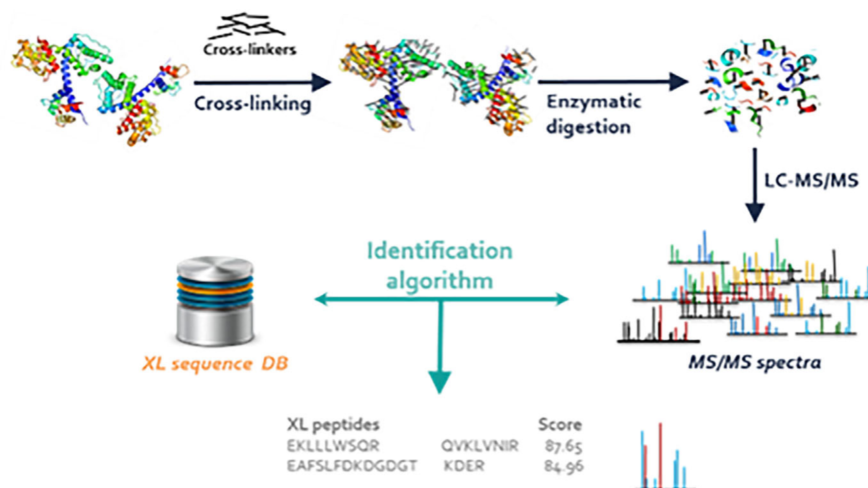

FIGURE 1 The experimental setup of an XL-MS experiment. The main difference between an XL-MS and a regular shotgun experiment is the first step: the actual cross-linking step. The subsequent steps—denaturation, enzymatic digestion and MS/MS analysis—are similar. The result is that MS/MS spectra from both single and cross-linked peptides are obtained. The subsequent identification of the cross-linked peptides is however much more challenging

Because unmodified linear peptides typically far outnumber cross-linked peptides in the sample, a number of strategies have been developed to enrich for cross-linked peptides. These strategies commonly exploit the aberrant chemical or physical properties that the cross-linkers confer onto the peptides. This can be achieved in one of three ways. The first approach utilizes trifunctional cross-linkers that contain an additional affinity tag (usually biotin or masked biotin).[11] The second approach uses strong cation exchange chromatography, and is based on the fact that unmodified tryptic peptides typically carry two positive charges (one positive charge on the N-terminus and another on the side chain of the C-terminal residue), while type 2 cross-linked peptides typically carry more than two charges as these contain two peptide N-termini and two basic C-terminal residues.[17] The third approach is based on size exclusion chromatography, as type 2 cross-linked peptides typically have a higher mass than unmodified linear peptides.[17,18] Another promising strategy is the use of cross-linkers that confer unique signatures in the MS step. With recent advances in cross-linker chemistry and the increasing resolution of mass spectrometers, this strategy has been gaining considerable support. The two major methods to identify cross-linked peptides at the MS level are: i) deuterated cross-linkers that give unique doublet signatures for cross-linked peptides in the MS1 spectra; or ii) CID-cleavable cross-linkers that create characteristic marker ions in fragmentation (MS/MS) spectra.

The third challenge in the identification of cross-linked peptides is related to fragmentation efficiency. In some cases, there are not enough fragment ions from each peptide present in the spectrum to identify both peptides confidently. This problem is most acute when peptides of very different lengths are linked and forms the base of the alpha and beta peptide terminology. The alpha peptide which is typically the longer peptide, is the peptide that shows superior fragmentation: it is well fragmented and the fragment ions have a high peak intensity. The beta peptide, the shorter peptide, fragments less well.[12,17,18]

MS/MS spectra of cross-linked peptides are in general much more complex than spectra of unmodified linear peptides. Cross-linked peptides tend to have higher precursor charges. Because these are composed of two covalently connected peptides, four (or more) charges can easily be accommodated. There are also many more fragment ions to consider because fragmentation can happen on any of the two peptides. This complexity increases further when the higher charge state of the precursor ions is taken into account, as this can result in fragment ions with higher charge states as well.

The last challenge is the enormously increased search space. In theory, each peptide with one or more cross-linker-modifiable residues can be linked to any other peptide with modifiable residues; in other words, a very large number of peptide-peptide combinations are possible from the theoretical digest of a sequence database, which results in an enormous increase of the search space. The number of possible peptide-to-peptide combinations is hence: $(n^2 + n)/2$ where n stands for the number of peptides (the n-square problem).[9] The n-square problem is nicely illustrated in Figure 3. To generate this plot, we downloaded all 20 130 human proteins from UniProtKB/SwissProt[19] (January 2017). These proteins were subsequently digested in silico by use of the compomics-utilities code library[20] with the following settings: trypsin as protease, two allowed missed cleavages, and minimum and maximum peptide lengths of respectively 4 and 20. From these 20 130 proteins, we randomly chose 200 proteins to calculate their number of possible cross-linkages based on DSS as cross-linker. From this simulation, one can see that, with these settings, the number of possible cross-linked peptides from 50 proteins already reaches the total number of single peptides that can be obtained from the whole human proteome. When possible cross-linked peptides from 100 proteins are considered, their total number is six times the number of single peptides obtained from the whole human proteome.

3 | ANALYSIS OF CROSS-LINKED PEPTIDE MASS SPECTRA

In an MS-based cross-linking experiment, residues are covalently linked with the aid of a cross-linker. There are about 100 cross-linkers described in literature and most of these are commercially available.[21] Despite this large diversity among cross-linkers, they all share the same basic design. A cross-linker is typically composed of two reactive functions which are separated by a spacer arm that usually consists of a single carbon chain. The two reactive functions covalently interact with the residues of a protein or complex. The most commonly used reactive function is an N-hydroxysuccinimide (NHS) ester which preferentially interacts with the primary amine of Lysine or the protein N-terminus.[22] Cross-linkers can either be homobifunctional, with two identical reactive functions that target identical functional groups on the protein or complex, or heterobifunctional, with two different reactive functions that target different functional groups on the protein or complex.[11]

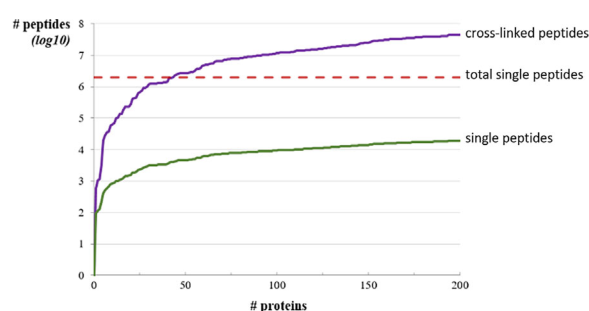

FIGURE 3 Illustration of the n-square problem. The number of peptides (y-axis) and proteins (x-axis) are plotted to illustrate the increase in search space. The number of possible cross-linked peptides is given in purple. The number of single tryptic peptides (two missed cleavages allowed) is given in green. The total number of single peptides from the human proteome is given in red. One can see that the number of possible cross-linked peptides from 50 proteins already reaches the number of total single peptides obtained from the whole human proteome. When the number of possible cross-linked peptides from 100 proteins is considered, this number is already six times the number of total single peptides obtained from the whole human proteome

The two reactive groups of the cross-linker are separated by a linker or spacer arm. The length of this spacer arm ranges from zero-length to about 40 Å and determines the maximal distance between two residues that can be covalently linked. It is the length of the spacer arm that determines the distance restraint imposed by the cross-linker. Zero-length cross-linkers, such as the carbodiimides (EDC and DCC), and N-hydroxysuccinimide (NHS) have no spacer arm and as such do not add atoms to the cross-linked species. This implies that the residues that are linked need to be within salt-bridge distance (less than or equal to 4 Å between the charged groups)[23] of each other. Zero-length cross-linkers hence capture direct contacts and not just residues that are in close proximity.[24]

Apart from the classification of cross-linkers as homo- or heterobifunctional, or by the length of their spacer arm, they can also be classified by the fragmentation behavior of the cross-linker, which results in intact and MS-cleavable cross-linkers. The former remain intact during fragmentation in the mass spectrometer, while the latter are designed to break apart in the mass spectrometer due to a labile bond that can easily be fragmented.[22] The spectra that are derived from each type of cross-linker are quite different and require a different type of interpretation. Because many researchers show a preference towards a given type of cross-linker, algorithms to identify linked peptides that are developed by these researches tend to be specific towards a single type of cross-linker. The next two sections therefore summarize the various algorithms that are available to identify cross-linked peptides obtained by each type of cross-linker.

3.1 | Identification of peptides linked by intact cross-linkers

Intact cross-linkers were the first cross-linkers to be used.[8] Peptides linked by an intact cross-linker stay together in the mass spectrometer. Therefore, the measured MS/MS spectrum is composed of fragment ions that originate from two peptides and is hence much more complex than the spectrum of a single peptide. The homobifunctional cross-linkers disuccinimidyl suberate (DSS) and bis(sulfosuccinimidyl)suberate (BS3) are the most commonly used intact cross-linkers. Both have two NHS esters as functional groups which interact with the primary amino group of Lysine or the protein N-terminus. The length of their spacer arm is 11.4 Å and due to the length of the side chain of Lysine (6.4 Å), DSS and BS3 have a distance constraint, ie, the maximum distance in space between two residues that a cross-linker can span, of 24.2 Å (as measured between the two Cα atoms).[25]

Because of the high complexity of MS/MS spectra that originate from cross-linked peptides, the established identification algorithms for single peptides such as Mascot[26] and X!Tandem[27] are not well-suited to identify cross-linked peptides. Manual assignment of cross-linked peptides to MS/MS spectra was therefore a common practice in the early days. One of the first approaches to computationally assign cross-linked peptides to MS/MS spectra was designed in such a way that already available search algorithms for single peptide identification could be used. This approach was based on the linearization of cross-linked peptide pairs.[28] All possible cross-linked peptides are introduced in the search database as linear sequences concatenated in pairs and this in all possible permutations. This linearized cross-linked database is then searched with traditional search algorithms and the mass of the cross-linker is introduced as a variable modification. However, in this approach, two linearized permutations derived from a pair of cross-linked peptides can provide the entire set of fragment ions only when considered together, which implies that each permutation results in some missing fragment ion types.[17] The strategy of linearized cross-linked peptides has later been used in several other dedicated cross-linked peptides identification algorithms, such as MSBridge[29] which is implemented in Protein Prospector[30] and pLink.[31]

Several tools apply a two-step approach for identification. During the first step, the peptide is identified based on its fragment ions which have the most intense peaks. This implies that it is typically the alpha peptide that is identified during this first step. The second step can then either rely on the identification of the second peptide, usually the beta peptide, by its mass as is the case in MassAI[32] or by looking for possible linkages on the first peptides as is applied by MXDB[33] and Kojak.[34] Xilmass[35] uses yet another approach to identify cross-linked peptides. The calculated theoretical spectra contain all possible fragment ions from a cross-linked peptide pair. Other algorithms typically do not build one single theoretical spectrum that contains all possible fragment ions of a cross-linked peptide pair. Instead, either individual theoretical spectra of each peptide in a cross-linked peptide pair are considered, or theoretical spectra generated from linearized peptides with the introduction of a variable modification.

Another way to identify cross-linked peptides relies on the usage of isotopically labeled cross-linkers. When the labelled cross-linkers are added in a 1:1 ratio (labeled vs non-labeled), the measured spectra contain characteristic mass doublets which can be used to identify spectra that originate from cross-linked peptides.[36] Both Hekate[36] and pLink[31] rely on this labeling strategy. A variant makes use of reporter ions; ions that originate from the cross-linker and have a characteristic footprint in the measured spectra. SIM-XL uses reporter ions to identify spectra that originate from cross-linked peptides.[37] Table 1 provides a short description of all the tools that are available to identify peptides linked by intact cross-linkers.

As mentioned before, a special type of intact cross-linkers are zero-length cross-linkers, which do not insert atoms during cross-linking. ZXMiner[24] was developed especially to identify peptides linked by zero-length cross-linkers. Another type of zero length cross-linkers which were extensively used in the past to probe protein structure are disulfide crosslinkers.[38] MassMatrix[39] was developed to identify either naturally occuring or chimeric (through engineered cysteine reisudes) disulfide bonds.

One important problem with the usage of intact cross-linkers is the enormously increased search space.[9] Tools to identify peptides linked by intact cross-linkers are therefore frequently limited to cross-linking experiments within a single protein or between a small number of proteins in a complex. A cross-linking experiment with an intact cross-linker on a whole proteome scale remains an essentially insurmountable challenge for most of the above described tools.

TABLE 1 Short description of the tools that are available to identify peptides linked by intact cross-linkers

Name	Short description	Url	Reference
MSBrigde	Identification based on linearization, implemented in protein prospector	http://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msbridgestandard	29
PeptideMap	Part of PROWL	http://prowl.rockefeller.edu/prowl/peptidemap.html	93
CLPM	Cross-link is treated as a modification	http://bioinformatics.ualr.edu/mbc/services/CLPM.html	94
Virtual-MSLAB	Based on matching of masses	Available upon request	95
MassAI	Past: CrossWork, one peptide is identified based on its fragment ions, the other by its mass	http://massai.dk/	32
StavroX	Comparison of precursor masses followed by calculation of b- and y-ions which are then matched against spectra	http://www.stavrox.com/	96
pLink	Labeling is optional, characteristic doublet peaks in MS1 spectra	http://pfind.ict.ac.cn/	31
Hekate	Labeling is necessary, characteristic mass doublets are used	http://andrewholding.com/research/hekate/	36
xQuest	For intact amine-reactive homobifunctional linkers	http://proteomics.ethz.ch/cgi-bin/xquest2_cgi/index.cgi	76,97
XLPM	Cross-linked peptide is treated as a modification	http://binf-app.host.ualr.edu/~mihir/cgi-bin/xlpm.cgi	98
CruX	Suite of algorithms, pFind Studio to identify cross-linked peptides	http://crux.ms/	99
anchorMS	Comparison of observed spectra with spectra for all possible cross-links	http://cbio.ufs.ac.za/AnchorMS/	100
MXDB	Two step approach, first identification of peptide with the best match then looking for all possible linkages	http://proteomics.ucsd.edu/software-tools/mxdb/	33
SIM-XL	Relies on reporter ions that originate from the cross-linker	http://patternlabforproteomics.org/sim-xl/	37
Kojak	Two step approach: first search for single peptides with modifications on linkable residues; the top scored spectra are then searched for cross-linked peptides	http://www.kojak-ms.org/	34
XLinkX	Mainly for MS-cleavable cross-linkers but can be used for intact cross-linkers	https://xlinkx.hecklab.com/	47
XL-Search	The scoring function estimates the joint posterior probability that both peptides are correctly identified	https://github.com/COL-IU/XLSearch	101
ECL	Cross correlation coefficient to measure similarity between theoretical and experimental spectra	http://bioinformatics.ust.hk/ecl.html	102
Xilmass	All possible fragment ions of the cross-linked peptides are in the theoretical spectrum	https://github.com/compomics/xilmass	35

These experiments can, however, be tackled with cleavable cross-linkers.

3.2 | Identification of peptides linked by MS-cleavable cross-linkers

In 2000, Bennett et al[40] were the first to describe the usage of a cleavable cross-linker. In their study, they made use of a thiol-cleavable cross-linker. This type of cleavable cross-linker relies on the comparison of peptide maps to identify cross-links: one peptide map that originates from the non-cleaved cross-linker and one peptide map from the cleaved cross-linker.[40] The usage of this type of cross-linkers does not, however, allow straightforward identification of cross-linked peptides and subsequently, of the cross-linked residues. This problem was solved by the introduction of MS-cleavable cross-linkers; cross-linkers that fragment within the mass spectrometer.

MS-cleavable cross-linkers have one or two labile bonds that can easily be fragmented, usually in collision induced dissocation (CID) conditions.[22] Peptides linked by MS-cleavable cross-linkers can therefore be observed both as cross-linked and individual peptides. This type of cross-linkers typically require a mass spectrometer with MS/MS/MS (MS3) capabilities. The MS1 spectra contain the intact cross-linked peptides. The MS/MS (MS2) spectra contain the single peptides which originate from CID-induced fragmentation of the cross-linker. The subsequent MS3 spectra contain the fragment ions of the individual peptides which allow for individual identification of the cross-linked peptides. The usage of MS-cleavable cross-linkers hence requires more advanced mass spectrometers. However, the latest mass spectrometers allow the observation of both the characteristic doublets of the cross-linker and the backbone fragments of the connected peptides on the MS2 level when BuUrBu is used as a cross-linker.[41]

BLinks[42] was the first algorithm that allowed the identification of peptides linked by MS-cleavable cross-linkers. BLinks relies on the usage of protein interaction reporters (PIRs) which were already introduced in 2005.[43] PIRs are a type of MS-cleavable cross-linkers which have two labile bonds in the spacer that surround a third function; an affinity tag for enrichment techniques. Fragmentation of these labile bonds results in the release of this MS reporter tag—specific constant neutral losses or characteristic fragment ions—which can easily be detected in the measured spectra and is as such a reporter for the presence of cross-linked peptides.[21] PIRs can have a relatively long spacer arm, up to 43 Å, but because of the high flexibility of the spacer arm, shorter distances can also be bridged.[22] BLinks uses both mass relationships and chromatographic information to identify linked peptides. A Mascot search[44] is subsequently used to identify the fragment ions from the cleaved peptides.[42]

MeroX[45] is another algorithm that is available to identify MS-cleavable linked peptides. It relies on finding signature peaks that are derived from MS-cleavable cross-linkers such as BuUrBu.[46] XLinkX[47] also identifies MS-cleavable linked peptides based on the identification of signature peaks but strongly depends on the experimental setup: CID to fragment the linked peptides followed by electron transfer dissociation (ETD) to fragment the individual peptides. Identification is a two-step proces. First, the precursor masses of the linked peptides are retrieved; then, spectra that contain at least one precursor mass pair are considered as potential cross-linked spectra, ie, all deduced cross-linked pairs undergo peptide-sequence analysis. The most recent version of XlinkX (XlinkX2) relies on finding peptide pairs.[48] Table 2 gives a short description of the tools that are available to identify peptides linked by MS-cleavable linkers.

3.3 | Quantification of cross-linked peptides

A cross-linking experiment allows scientists to explore the possible conformations, and intra and inter molecular interactions of a protein. Therefore, quantification of the relative abundance of reporter cross-linked peptides from each unique conformation or interaction state would allow the analysis of the distribution of these different conformations and as such provides insights into the dynamics of the protein. Recent applications of this methodology mostly involved either labeling of the cross-linker or the interacting proteins (traditional SILAC based methodologies) with heavy and light isotopes.[49–52] However, recently a label-free quantification-linkage methodology has been reported, in which an isotope labeling step is no longer required.[49] Only two tools are available for the quantification of a cross-linking experiment: XiQ[53] and xTract.[49] As in bottom-up proteomics, both tools rely on the observation that the intensity of the peak (signal intensity) is proportional to the concentration of the peptide. XiQ relies on the use of labelled cross-linkers. This implies that an additional labeling step is unnecessary. XiQ first extracts the peak volumes of the precursor ion and that of the mass doublet signal. These peak volumes are then used to calculate the heavy over light cross-linked peptide ratio.[53] Based on XiQ, a novel version of MaxQuant was introduced[54] that enables the quantification of cross-linked peptides. It implements a "quantification only" mode. This mode allows

TABLE 2 Short description of the software tools that are available to identify peptides linked by MS-cleavable cross-linkers

Name	Short description	Url	Reference
BLinks	Relies on PIRs for identification, and Mascot to identify the individual peptides	http://brucelab.gs.washington.edu/software.html	42
MeroX	Relies on signature peaks from MS-cleavable cross-linkers for identification	http://www.stavrox.com/	45
XLinkX	Requires CID followed by ETD as experimental setup	https://sourceforge.net/projects/xlinkx/	47

quantification independent of the identification. The intensity ratio is calculated as in XiQ, based on the peaks of the mass doublets.[52] xTract allows quantification based on both labelled and label-free data.[49] Overall, quantitative cross-linking experiments have been shown to help understand the dynamics of protein complexes and interactions, but the exploration of computational solutions to analyze these data has only just started.

3.4 | The false discovery rate in cross-linked peptide identification

In shotgun proteomics, there is always a degree of uncertainty associated with the accuracy of the peptide-to-spectrum matches (PSMs). This is also true in XL-MS experiments. In shotgun proteomics, this overall reliability of the obtained PSMs is assessed by the target-decoy approach[55] in which a target sequence is either reversed or shuffled to generate a decoy sequence. Searching the experimental spectra against a concatenated target-decoy database then allows the computation of the false discovery rate (FDR).[56–58] Several groups adapted this traditional target-decoy strategy to calculate the FDR in XL-MS experiments. This adaptation is needed because every PSM is a match to two peptides which each have their own probability of being false. This means that, to calculate the FDR, three possibilities need to be taken into account: decoy-decoy and target-target combinations but also target-decoy combinations. xProphet takes all these possible target and decoy combinations into account to calculate the FDR in XL-MS experiments.[59]

In cross-linking, the FDR is typically calculated for peptide pairs. But from an experimental point of view, the focus is on residues that are covalently linked. This implies that the FDR can be approached from different levels: PSM, peptide pairs, residues pairs and even protein pairs. Fisher and Rappsilber[60] calculated that 5% FDR at the level of PSMs leads to 5.8% FDR at the level of peptide pairs and already 8.4% FDR at the level of residue pairs. This illustrates that the FDR at the level of the PSMs is not a good guide for the FDR at the level of the residue pairs. They therefore propose to prefilter the FDR at the level of the PSM and peptide pair and report the FDR at the level of the residue pair.[60]

There are several causes for wrongly assigned PSMs. These are nicely described by Iacobucci and Sinz[61] and briefly presented here. One cause is the presence of isobaric cross-linked peptides: cross-links that occur between peptides that are consecutive in the protein sequence are isobaric to dead-end peptides that involve the same amino acid sequence. It is therefore important to look at characteristic fragment ions from both cross-linked peptides. The use of MS-cleavable cross-linkers can also overcome this problem.[61] Another problem is related to incomplete fragementation of both cross-linked products as discussed earlier. One peptide (alpha) is often thoroughly sequenced while the other peptide (beta) is often only little or not fragmented and hence ambiguous. It is therefore advised to discard cross-links if both peptides are not thoroughly sequenced.[61] Cross-linking experiments also require high mass accuracy and this both at the MS1 and MS2 level where a 5 ppm mass tolerance for the precursor ions and a 10 ppm mass tolerance for the product ions is suggested.[61] Wrongly assigned PSMs can also be due to a high number of unassigned peaks or low signal-to-noise ratio. Iacobucci and Sinz[61] therefore advise to discard spectra with low signal-to-noise ratios and to accept cross-linked peptides only if the majority of the fragment ions can be assigned.

4 | VISUALIZATION OF IDENTIFIED CROSS-LINKED PEPTIDES

The goal of a cross-linking experiment is to obtain information about the structure of a protein or protein complex, the dynamics within a protein, or to gain knowledge about interaction partners. Hence, once the measured cross-links are identified, these need to be interpreted further because a list of cross-linked peptides by itself does not satisfy this purpose. Visualization of the outcome of a cross-linking experiment is therefore an important step in the analysis of the data. Several tools have been developed to allow such visualization of cross-linking data, and a short description of each is given in Table 3.

The simplest way to visualize the outcome of a cross-linking experiment is by mapping the identified cross-linked sites on the 3D structure of the protein or complex. XWalk[62] performs this mapping of measured cross-links on the available structure, and can also calculate the length of the shortest path through the solvent between two cross-linked residues without penetration of the protein surface. This distance is termed the solvent accessible surface distance (SASD) and is typically larger than the Euclidean distance between the two linked residues. Combined with a known linker length, knowledge of this distance provides information about which measured linkages violate the SASD and as such might be an indication of deviations from the existing structure. XWalk can also be used to predict possible cross-links: when a structure is available, the SASD between all possible linker residues can be calculated. Xlink Analyzer[16] is another tool that allows the visualization of cross-linked residues on available protein structures (Figure 4A). Xlink Analyzer is available as a plugin for the UCSF Chimera structure visualization software.[63] Like XWalk, it displays the identified cross-links on the protein structure and detects cross-links that violate the distance restraints of the cross-linker. Thanks to Xlink Analyzer's integration into UCSF Chimera, it can also visualize several other properties of the protein of interest (ie, solvent accessibility and electrostatic potential, . . .). Moreover, mono-links can also be displayed and predicted to assess the surface accessible residues.

Cross-links sometimes violate distance restraints; given the limited reach of the linker, their link can be seemingly incompatible with the existing structure of a protein. This, however, does not mean that the identified cross-link is wrong. The cross-link might originate from an alternative conformation of the protein. DynaXL[64] assesses conformational fluctuations of multidomain proteins. It calculates possible alternative conformations based on identified cross-links and rigid body/torsion angle refinement and as such uses the supposed wrong distances to optimize the possible conformation of the protein complex. XL-MOD[65] goes one step further. It uses the distance restraints of the

TABLE 3 Short description of the tools that are available to visualize the results from a cross-linking experiment

Name	Short description	2D	3D	Url	Reference
ProXL	Web-application to analyze, share and visualize cross-linking data	Yes	Yes	https://github.com/yeastrc/proxl-web-app	66
xiNet	Representation of cross-links in node-like diagrams	Yes	No	http://crosslinkviewer.org/	74
Xlink-analyzer	Plugin for UCSF Chimera, to display cross-links on protein structures	No	Yes	http://www.beck.embl.de/XlinkAnalyzer.html	16
XLinkDB	Network analysis based on cross-linking data	Yes	Yes	http://xlinkdb.gs.washington.edu/	67
xVis	Web-application to visualize and interpret cross-links	Yes	No	https://xvis.genzentrum.lmu.de/	73
XWalk	Prediction and validation of cross-links	No	Yes	http://www.xwalk.org/	62
DynaXL	Visualization of conformational fluctuations of multidomain proteins	No	Yes	http://www.tanglab.org/resources/dynaxl	64
XL-MOD	Automated modeling of large multi-subunit protein complexes	No	Yes	http://aria.pasteur.fr/supplementary-data/x-links	65

cross-linked residues to automatically model large multi-subunit protein complexes. To resolve conflicting data, that is identified cross-links that are too distant based on linker length and known structures, subunits are allowed flexibility during conformational sampling.

ProXL[66] is a web-based application that allows the analysis and visualization of cross-linked data. In contrast to XWalk and Xlink Analyzer, ProXL also visualizes cross-link data in 2D and represents protein sequences as bars with linkages between these. Moreover, ProXL allows to store and share cross-linking data between researchers. XLinkDB[67] also allows to store and share cross-linking projects. It also gives the option to map cross-linking data on to structures that are generated by homology modeling with Modeller,[68] or on to complexes that are obtained by docking with PatchDock.[69] Both tools are integrated via the Integrative Modeling Platform (IMP).[70] XLinkDB also has the option to perform network analysis.

At the time of writing, the number of protein structures (131 205 available in the PDB[71] as of June 2017; www.pdb.org) remains small compared to the number of annotated protein sequences (87 846 192 sequences in UniprotKB[72] (June 2017)). Therefore, it is highly likely that no structure is available to map identified cross-linked peptides on. This is especially true when one wants to map data on protein complexes. XLinkDB therefore implemented modeling and docking modules. Still, not all structures of proteins or complexes can be obtained with these tools. It is therefore also very important to visualize cross-linking data on protein sequences. xVis[73] and xiNet[74] allow such visualization for proteins without any structural information. In xVis, proteins and complexes are represented as circular plots, bar plots or network diagrams (Figure 4C-E). xVis is also linked to InterPro[75] which allows domain information to be mapped on the protein. However, xVis supports the visualization of cross-link results obtained from any search engine provided that the input data is in the xQuest format.[76] xiNet presents the cross-linking data in node-like diagrams where each node represents a protein (Figure 4B), and is independent of the algorithm that was used to identify the cross-linked peptides.[74]

5 | DATA MANAGEMENT AND DISSEMINATION OF IDENTIFIED CROSS-LINKED PEPTIDES

Once the data have been processed and interpreted, the results, the data, and the metadata information need to be stored and disseminated. In the past, it was common practice to store data in-house only, if at all. But because scientists are increasingly appreciating the value of Open Science,[77,78] it has become more popular (or even mandatory) to make source data as well as accompanying experimental metadata public, alongside the results and conclusions typically reported in papers. Various platforms already exist to share protein structure data: the longest running such repository is the PDB,[71] which contains protein structures that were determined by X-ray crystallography, NMR or high resolution electron microscopy. And, as mentioned above, XLinkDB[67] and ProXL[66] provide online databases for cross-linking experiments.

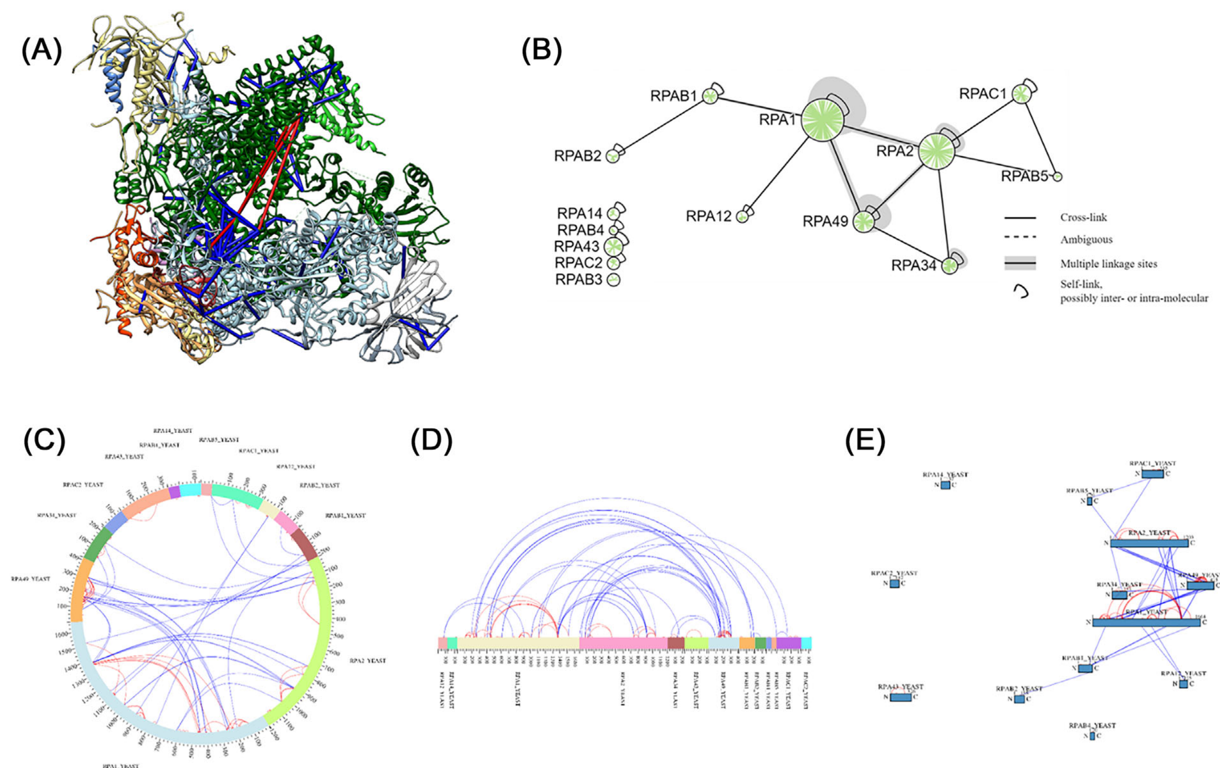

FIGURE 4 Identified cross-links can be visualized in several ways. A, When the structure of the protein or complex is available, the identified cross-links can be visualized on the structure itself. Here we show the identified cross-links of RNA polymerase I (pdb-entry 4C3H103) with the aid of Xlink Analyzer. Blue cross-links satisfy the structural constraints while red cross-links violate the distance threshold. These are an indication for flexibility. B, xiNet represents the cross-linking data in node-like diagrams. With xVis, the data can be represented in a circular plot (C), a bar plot (D) or in a network plot (E). The different types of visualization are illustrated on RNA polymerase I of Saccharomyces cerevisiae, a dataset that is available through Xlink Analyzer[16]

At the moment, XLinkDB contains several datasets of large-scale cross-linking studies together with the structural and docking predictions made from these experiments.[67] However, only the outcome of the experiments is stored in XLinkDB, and not the source data. It is therefore impossible to link back to the spectra from which the peptides were identified. ProXL, another web application that allows researchers to store, visualize and share their cross-linking data,[66] does store these source data. However, public availability is not standard in ProXL; instead users need to be explicitly given access to the data by the scientist that submitted the data.

For shotgun proteomics, several repositories are available to store raw data, metadata and results. PRIDE[79] and PeptideAtlas[80] are the most prominent repositories, and these two formed the foundation of the ProteomeXchange consortium that provides a common framework for user-friendly data deposition and exchange of mass spectrometry based proteomics data.[81] MassIVE (http://massive.ucsd.edu/ProteoSAFe/static/massive.jsp) and jPostrepo[82] have since joined ProteomeXchange. ProteomeXchange has recently also explicitly started to support storage of, and access to, data from cross-linking experiments.

An important aspect of data management and dissemination is standardization. Large efforts have been invested to standardize data from shotgun experiments. However, standardization in XL-MS is lagging behind. Each cross-linked peptide identification algorithm has its own way to report the obtained identifications. As a first attempt to standardization, an extension of the older pepXML format was published as a standard format in order to support the identification of cross-linked peptides.[83] Meanwhile, the HUPO-PSI group (https://github.com/HUPO-PSI/mzIdentML/, Accession Date: March 28, 2017) has released support for cross-linking data in version 1.2 of its mzIdentML standard.[84]

6 | CONCLUSIONS AND OUTLOOK

Rather than relying on a single approach to unravel protein structures, these days, a combination of several complementary approaches is often used to also obtain insight into protein dynamics and complex formation. Cross-linking analyzed by mass spectrometry has become an important actor in this proces. This is illustrated by the plentitude of published studies in which cross-linking was used to analyse protein complexes such as the study of the interaction partners and dynamics of the cannabinoid receptors,[85] the investigation of the interaction network of human protein kinase D2,[86] analysis of Ca^{2+}-induced changes in calreticulin,[87] unravelling of the architecture of the human polycomb respressive complex 2,[88] and many others.

Dispite the advances in cross-linking, there are only a limited number of studies published which tackle proteome-wide interaction

studies through cross-linking. Examples are the elucidation of the murine mitochondrial protein interactome,[89] mapping of the disulfide proteome,[90] analysis of HeLa cells[47] and *Shewanella oneidensis*.[91] This low number finds its origin in the challenges met when analysing cross-linked peptides by mass spectrometry. When performing a cross-linking experiment on a whole proteome, there are only a very low amount of cross-linked peptides compared to unmodified peptides. It is also much more difficult to reliably identify cross-linked peptides, especially when one peptide is much shorter than the other; the fragmentation efficiency of the latter is usually inadequate[18] as described above. Moreover, even when the shorter peptide produces enough fragment ions for identification, it is much more difficult to assign this peptide unambiguously as in the protein database, there might be several proteins that contain this peptide; a challenge which is of less significance when a cross-linking experiment is performed on a purified protein complex.[92]

Although XL-MS is a relatively underdeveloped field compared to traditional mass spectrometry-based shotgun experiments, many computational tools are already available to identify and further interprete cross-linked peptides. Due to a rapid increase in popularity of XL-MS, this number continues to increase alongside new developments in instrumentation and experimental approaches. In this review, we gave an overview of the available tools. Moreover, a detailed and up-to-date list of the described algorithms can be found at: http://iomics.ugent.be/xltools/. This list not only sums the different algorithms that are available but also provides information about the availability, some tools are publically available while others require a license, and ease to use, some tools have a web interface, others a GUI while still others are only available as command line tool.

Because most of these tools can be quite specific towards one type of cross-linker or experimental setup,[11] the choice of an appropriate tool can be confusing for researchers. With this review, we therefore hope to have provided the reader a helping hand in their choice for an algorithm that meets their needs.

ACKNOWLEDGMENTS

This work was supported by grants (to AE): RiMembR (Vlaanderen Onderzoeksprojecten; #G0C6814N; FWO); T3RecS (#G002516N; FWO) and DIP-BiD (#AKUL/15/40 – G0H2116N; Hercules/FWO) and (to JR): R&D Pilot project-2015 [Instruct, part of the European Strategy Forum on Research Infrastructures (ESFRI) and supported by national member subscriptions]. Concerted Research Action (BOF12/GOA/014) from Ghent University. E.V. is a postdoctoral research fellow of the Research Foundation Flanders.

REFERENCES

1. Tompa P. The interplay between structure and function in intrinsically unstructured proteins. *FEBS Lett*. 2005;579: 3346–3354.
2. Rudiger S, Freund SMV, Veprintsev DB, Fersht AR. CRINEPT-TROSY NMR reveals p53 core domain bound in an unfolded form to the chaperone Hsp90. *Proc Natl Acad Sci USA*. 2002;99: 11085–11090.
3. Kaltashov I, Bobst C, Abzalimov R. H/D exchange and mass spectrometry in the studies of protein conformation and dynamics: is there a need for a top-down approach? *Anal Chem*. 2009;81: 7892–7899.
4. Vandermarliere E, Stes E, Gevaert K, Martens L. Resolution of protein structure by mass spectrometry. *Mass Spectrom Rev*. 2106;35: 653–665.
5. Lasker K, Forster F, Bohn S, et al. Molecular architecture of the 26S proteasome holocomplex determined by an integrative approach. *Proc Natl Acad Sci U S A*. 2012;109:1380–1387.
6. Ward A, Sali A, Wilson I. Integrative structural biology. *Sci Rep*. 2013;339:913–915.
7. Benda C, Ebert J, Scheltema RA, et al. Structural model of a CRISPR RNA-silencing complex reveals the RNA-target cleavage activity in Cmr4. *Mol Cell*. 2014;56:43–54.
8. Fasold H, Klappenberger J, Meyer C, Remold H. Bifunctional reagents for the crosslinking of proteins. *Angew Chem Int Ed Engl*. 1971;10:795–801.
9. Rappsilber J. The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. *J Struct Biol*. 2011;173:530–540.
10. Sinz A. Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions. *Mass Spectrom Rev*. 2006;25:663–682.
11. Sinz A. Chemical cross-linking and mass spectrometry for mapping three-dimensional structures of proteins and protein complexes. *J Mass Spectrom*. 2003;38:1225–1237.
12. Schilling B, Row RH, Gibson BW, Guo X, Young MM. MS2Assign, automated assignment and nomenclature of tandem mass spectra of chemically crosslinked peptides. *J Am Soc Mass Spectrom*. 2003;14: 834–850.
13. Tosi A, Haas C, Herzog F, et al. XStructure and subunit topology of the INO80 chromatin remodeler and its nucleosome complex. *Cell*. 2013;154:1207–1219.
14. Bui KH, Von Appen A, Diguilio AL, et al. Integrated structural analysis of the human nuclear pore complex scaffold. *Cell*. 2013;155: 1233–1243.
15. Kahraman A, Herzog F, Leitner A, et al. Cross-link guided molecular modeling with ROSETTA. *PLoS ONE*. 2013;8:e73411.
16. Kosinski J, von Appen A, Ori A, et al. Xlink analyzer: software for analysis and visualization of cross-linking data in the context of three-dimensional structures. *J Struct Biol*. 2015; 189:177–183.
17. Leitner A, Walzthoeni T, Kahraman A, et al. Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics. *Mol Cell Proteomics*. 2010;9: 1634–1649.
18. Giese SH, Fischer L, Rappsilber J. A study into the collision-induced dissociation (CID) behavior of cross-linked peptides. *Mol Cell Proteomics*. 2016;15:1094–1104.
19. Bateman A, Martin MJ, O'Donovan C, et al. UniProt: a hub for protein information. *Nucleic Acids Res*. 2015;43:D204–D212.
20. Barsnes H, Vaudel M, Colaert N, et al. Compomics-utilities: an open-source Java library for computational proteomics. *BMC Bioinformatics*. 2011;12:70.
21. Paramelle D, Miralles G, Subra G, Martinez J. Chemical cross-linkers for protein structure studies by mass spectrometry. *Proteomics2*. 2013;13:438–456.
22. Sinz A. Divide and conquer: cleavable cross-linkers to study protein conformation and protein-protein interactions. *Anal Bioanal Chem*. 2017;409:33–44.
23. Barlow D, Thornton J. Ion-pairs in proteins. *J Mol Biol*. 1983;168: 867–885.
24. Rivera-Santiago R, Sriswasdi S, Harper S, Speicher D. Probing structures of large protein complexes using zero-length cross-linking. *Methods*. 2015;89:99–111.

25. Merkley ED, Rysavy S, Kahraman A, et al. Distance restraints from crosslinking mass spectrometry: mining a molecular dynamics simulation database to evaluate lysine-lysine distances. *Protein Sci*. 2014;23:747–759.
26. Perkins D, Pappin D, Creasy D, Cottrell J. Probability-based protein identification by searching databases using mass spectrometry data. *Electrophoresis*. 1999;20:3551–3567.
27. Craig R, Beavis R. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics*. 2004;20:1466–1467.
28. Maiolica A, Cittaro D, Borsotti D, et al. Structural analysis of multiprotein complexes by cross-linking, mass spectrometry, and database searching. *Mol Cell Proteomics*. 2007;6:2200–2211.
29. Chu F, Shan S, Moustakas DT, et al. Unraveling the interface of signal recognition particle and its receptor by using chemical cross-linking and tandem mass spectrometry. *Proc Natl Acad Sci USA*. 2004;101:16454–16459.
30. Trnka MJ, Baker PR, Robinson PJJ, Burlingame AL, Chalkley RJ. Matching cross-linked peptide spectra: only as good as the worse identification. *Mol Cell Proteomics*. 2014;13:420–434.
31. Yang B, Wu Y, Zhu M, et al. Identification of cross-linked peptides from complex samples. *Nat Methods*. 2012;9:904–906.
32. Rasmussen MI, Refsgaard JC, Peng L, Houen G, Højrup P. Cross-Work: software-assisted identification of cross-linked peptides. *J Proteomics*. 2011;74:1871–1883.
33. Wang J, Anania VG, Knott J, et al. Combinatorial approach for large-scale identification of linked peptides from tandem mass spectrometry spectra. *Mol Cell Proteomics*. 2014;13:1128–1136.
34. Hoopmann MR, Zelter A, Johnson RS, et al. Kojak: efficient analysis of chemically cross-linked protein complexes HHS public access. *J Proteome Res*. 2015;14:2190–2198.
35. Yllmaz Ş, Drepper F, Hulstaert N, et al. Xilmass: a new approach toward the identification of cross-linked peptides. *Anal Chem*. 2016;88:9949–9957.
36. Holding A, Lamers M, Stephens E, Skehel J. Hekate: software suite for the mass spectrometric analysis and three-dimensional visualization of cross-linked protein samples. *J Proteome Res*. 2013;12:5923–5933.
37. Lima DB, de Lima TB, Balbuena TS, et al. SIM-XL: a powerful and user-friendly tool for peptide cross-linking analysis. *J Proteomics*. 2014;129:51–55.
38. Pakula A, Simon M. Determination of transmembrane protein structure by disulfide cross-linking: the *Escherichia coli* tar receptor. *Proc Natl Acad Sci U S A*. 1992;89:4144–4148.
39. Xu H, Zhang L, Freitas MA. Identification and characterization of disulfide bonds in proteins and peptides from tandem MS data by use of the MassMatrix MS/MS search engine. *J Proteome Res*. 2008;7:138–144.
40. Bennett KL, Kussmann M, Mikkelsen M, et al. Chemical cross-linking with thiol-cleavable reagents combined with differential mass spectrometric peptide mapping-A novel approach to assess inter-molecular protein contacts. *Protein Sci*. 2000;9:1503–1518.
41. Arlt C, Götze M, Ihling CH, et al. Integrated workflow for structural proteomics studies based on cross-linking/mass spectrometry with an MS/MS cleavable cross-linker. *Anal Chem*. 2016;88:7930–7937.
42. Hoopmann MR, Weisbrod C, Bruce J. Improved strategies for rapid identification of chemically cross-linked peptides using protein interaction reporter technology. *J Proteome Res*. 2010;9:6323–6333.
43. Tang X, Munske GR, Siems WF, Bruce JE. Mass spectrometry identifiable cross-linking strategy for studying protein-protein interactions. *Anal Chem*. 2005;77:311–318.
44. Perkins D, Pappin D, Creasy D, Cottrell J. Probability-bases protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*. 1999;20:3551–3567.
45. Götze M, Pettelkau J, Fritzsche R, et al. Automated assignment of MS/MS cleavable cross-links in protein 3d-structure analysis. *J Am Soc Mass Spectrom*. 2014;26:83–97.
46. Müller MQ, Dreiocker F, Ihling CH, Schäfer M, Sinz A. Cleavable cross-linker for protein structure analysis: reliable identification of cross-linking products by tandem MS. *Anal Chem*. 2010;82:6958–6968.
47. Liu F, Rijkers DTS, Post H, Heck AJR. Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry. *Nat Methods*. 2015;12:1179–1184.
48. Liu F, Lössl P, Scheltema R, Viner R, Heck AJR. Optimized fragmentation schemes and data analysis strategies for proteome-wide cross-link identification. *Nat Commun*. 2017;8:15473.
49. Walzthoeni T, Joachimiak LA, Rosenberger G, et al. XTract: software for characterizing conformational changes of protein complexes by quantitative cross-linking mass spectrometry. *Nat Methods*. 2015;12:1185–1190.
50. Yu C, Huszagh A, Viner R, et al. Developing a multiplexed quantitative cross-linking mass spectrometry platform for comparative structural analysis of protein complexes. *Anal Chem*. 2016;88:10301–10308.
51. Zhong X, Navare AT, Chavez JD, et al. Large-scale and targeted quantitative cross-linking MS using isotope-labeled protein interaction reporter (PIR) cross-linkers. *J Proteome Res*. 2017;16:720–727.
52. Chen ZA, Fischer L, Cox J, Rappsilber J. Quantitative cross-linking/mass spectrometry using isotope-labeled cross-linkers and MaxQuant. *Mol Cell Proteomics*. 2016;15:2769–2778.
53. Fischer L, Chen ZA, Rappsilber J. Quantitative cross-linking/mass spectrometry using isotope-labelled cross-linkers. *J Proteomics*. 2013;88:120–128.
54. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol*. 2008;26:1367–1372.
55. Elias J, Gibbons F, King O, Roth F, Gygi S. Intensity-based protein identification by machine learning from a library of tandem mass spectra. *Nat Biotechnol*. 2004;22:214–219.
56. Käll L, Storey JD, Maccoss MJ, Noble WS. Posterior error probabilities and false discovery rates: two sides of the same coin posterior error probabilities and false discovery rates: two sides of the same coin. *J Proteome Res*. 2008;7:40–44.
57. Nesvizhskii AI. A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics. *J Proteomics*. 2010;73:2092–2123.
58. Colaert N, Degroeve S, Helsens K, Martens L. Analysis of the resolution limitations of peptide identification algorithms. *J Proteome Res*. 2011;10:5555–5561.
59. Walzthoeni T, Claassen M, Leitner A, et al. False discovery rate estimation for cross-linked peptides identifies by mass spectrometry. *Nat Methods*. 2012;9:901–903.
60. Fisher L, Rappsilber J. Quirks or error estimation in cross-linking/mass spectrometry. *Anal Chem*. 2017;89:3829–3833.
61. Iacobucci C, Sinz A. To be or not to be? Five guidelines to avoid misassignments in cross-linking/mass spectrometry. *Anal Chem*. 2017;89:7832–7835.
62. Kahraman A, Malmström L, Aebersold R. Xwalk: computing and visualizing distances in cross-linking experiments. *Bioinformatics*. 2011;27:2163–2164.
63. Pettersen EF, Goddard TD, Huang CC, et al. UCSF chimera − a visualization system for exploratory research and analysis. *J Comput Chem*. 2004;25:1605–1612.
64. Ding Y-H, Gong Z, Dong X, et al. Modeling protein excited-state structures from "over-length" chemical cross-links. *J Biol Chem*. 2016;292:jbc.M116.761841.
65. Ferber M, Kosinski J, Ori A, et al. Automated structure modeling of large protein assemblies using crosslinks as distance restraints. *Nat Methods*. 2016;13:515–520.
66. Riffle M, Jaschob D, Zelter A, Davis TN. ProXL (protein cross-linking database): a platform for analysis, visualization, and sharing of

protein cross-linking mass spectrometry data. *J Proteome Res*. 2016;15:2863–2870.
67. Schweppe DK, Zheng C, Chavez J, et al. XLinkDB 2.0: integrated, large-scale structural analysis of protein crosslinking data. *Bioinformatics*. 2016;32:2716–2718.
68. Fiser A, Sali A. Modeller: generation and refinement of homology-based protein structure models. *Methods Enzym*. 2003;374: 461–491.
69. Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson H. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res*. 2005;22:W363–W367.
70. Russel D, Lasker K, Webb B, et al. Putting the pieces together: integrative modeling platform software for structure determination of macromolecular assemblies. *PLoS Biol*. 2012;10:e1001244.
71. Berman HM, Westbrook J, Feng Z, et al. The protein data bank. *Nucleic Acids Res*. 2000;28:235–242.
72. The Uniprot Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res*. 2017;45:D158–D169.
73. Grimm M, Zimniak T, Kahraman A, Herzog F. XVis: a web server for the schematic visualization and interpretation of crosslink-derived spatial restraints. *Nucleic Acids Res*. 2015;43:W362–W369.
74. Combe CW, Fischer L, Rappsilber J. XiNET: cross-link network maps with residue resolution. *Mol Cell Proteomics*. 2015;14:1137–1147.
75. Finn RD, Attwood TK, Babbitt PC, et al. InterPro in 2017-beyond protein family and domain annotations. *Nucleic Acids Res*. 2017;45: D190–D199.
76. Leitner A, Walzthoeni T, Aebersold R. Lysine-specific chemical cross-linking of protein complexes and identification of cross-linking sites using LC-MS/MS and xQuest/xProphet software pipeline. *Nat Protoc*. 2014;9:120–137.
77. Tennant JP, Waldner F, Jacques DC, et al. The academic, economic and societal impacts of open access: an evidence-based review. *F1000Research*. 2016;5:632.
78. Masuzzo P, Martens L. Do you speak open science? Resources and tips to learn the language. *PeerJ Preprints*. 2017;5:e2689v1.
79. Martens L, Hermjakob H, Jones P, et al. PRIDE: the proteomics identifications database. *Proteomics*. 2005;5:3537–3545.
80. Deutsch EW, Eng J, Zhang H, et al. Human plasma PeptideAtlas. *Proteomics*. 2005;5:3497–3500.
81. Vizcaíno JA, Deutsch EW, Wang R, et al. ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat Biotechnol*. 2014;32:223–226.
82. Okuda S, Watanabe Y, Moriya Y, et al. JPOSTrepo: an international standard data repository for proteomes. *Nucleic Acids Res*. 2017;45: D1107–D1111.
83. Hoopmann MR, Mendoza L, Deutsch EW, Shteynberg D, Moritz RL. An open data format for visualization and analysis of cross-linked mass spectrometry results. *J Am Soc Mass Spectrom*. 2016;27: 1728–1734.
84. Vizcaíno JA, Mayer G, Perkins SR, et al. The mzIdentML data standard version 1.2, supporting advances in proteome informatics. *Mol Cell Proteomics*. 2017;16:1275–1285.
85. Song Z, Laun A, Cai J. Mass spectrometry analysis of human CB2 cannabinoid receptor and its associated proteins. *Methods Enzym*. 2017;593:371–386.
86. Häupl B, Ihling C, Sinz A. Protein interaction network of human protein kinase D2 revealed by chemical cross-linking/mass spectrometry. *J Proteome Res*. 2016;15:3686–3699.
87. Boelt S, Norn C, Rasmussen MI, et al. Mapping the Ca(2+) induced structural change in calreticulin. *J Proteomics*. 2016;142: 138–148.
88. Ciferri C, Lander G, Maiolica A, et al. Molecular architecture of human polycomb repressive complex 2. *Elife*. 2012;5:e00005.
89. Schweppe DK, Chavez JD, Lee C, et al. Mitochondrial protein interactome elucidated by chemical cross-linking mass spectrometry. *Proc Natl Acad Sci U S A*. 2017;114:1732–1737.
90. Lu S, Fan S-B, Yang B, et al. Mapping native disulfide bonds at a proteome scale. *Nat Methods*. 2015;12:329–331.
91. Zhang H, Tang X, Munske G, et al. Identification of protein-protein interactions and topologies in living cells with chemical cross-linking and mass spectrometry. *Mol Cell Proteomics*. 2009;8: 409–420.
92. Barysz H, Malmstroem J. Development of large-scale cross-linking mass spectrometry. *Mol Cell Proteomics*. 2017.
93. Beavis R, Fenyö D. Finding protein sequences using PROWL. *Curr Protoc Bioinforma*. 2004:chapter 13, unit 13.2.
94. Tang Y, Chen Y, Lichti C, et al. CLPM: a cross-linked peptide mapping algorithm for mass spectrometric analysis. *BMC Bioinformatics*. 2005;15:S9.
95. de Koning L, Kasper P, Back JW, et al. Computer-assisted mass spectrometric analysis of naturally occurring and artificially introduced cross-links in proteins and protein complexes. *FEBS J*. 2006;273:281–291.
96. Götze M, Pettelkau J, Schaks S, et al. StavroX-a software for analyzing crosslinked products in protein interaction studies. *J Am Soc Mass Spectrom*. 2012;23:76–87.
97. Rinner O, Seebacher J, Walzthoeni T, et al. Identification of cross-linked peptides from large sequence databases. *Nat Methods*. 2008;5:315–318.
98. Jaiswal M, Crabtree N, Bauer M, et al. XLPM: efficient algorithm for the analysis of protein-protein contacts using chemical cross-linking mass spectrometry. *BMC Bioinformatics*. 2014;15:S16.
99. McIlwain S, Tamura K, Kertesz-Farkas A, et al. Crux: rapid open source protein tande mass spectrometry analysis. *J Proteome Res*. 2014;13:4488–4491.
100. Mayne S, Patterton H. AnchorMS: a bioinformatics tool to derive structural information from the mass spectra of cross-linked protein complexes. *Bioinformatics*. 2014;30:125–126.
101. Ji C, Li S, Reilly J, Radivojac P, Tang H. XLSearch: a probabilistic database search algorithm for identifying cross-linked peptides. *J Proteome Res*. 2016;15:1830–1841.
102. Yu F, Li N, Yu W. ECL: an exhaustive search tool for the identification of cross-linked peptides using whole database. *BMC Bioinformatics*. 2016;17:217.

How to cite this article: Yılmaz Ş, Shiferaw GA, Rayo J, Economou A, Martens L, Vandermarliere E. Cross-linked peptide identification: A computational forest of algorithms. *Mass Spec Rev*. 2018;1–12. https://doi.org/10.1002/mas.21559