Cross-linking mass spectrometry

Citation for published version:
O'Reilly, FJ & Rappsilber, J 2018, 'Cross-linking mass spectrometry: methods and applications in structural, molecular and systems biology', *Nature Structural & Molecular Biology*. https://doi.org/10.1038/s41594-018-0147-0

Digital Object Identifier (DOI):
10.1038/s41594-018-0147-0

Published In:
Nature Structural & Molecular Biology

Cross-linking/mass spectrometry: methods and applications to structural, molecular and systems biology

Francis J O'Reilly[1,2] and Juri Rappsilber*[1,2]
[1]Bioanalytics, Institute of Biotechnology, Technische Universität Berlin, 13355 Berlin, Germany
[2]Wellcome Centre for Cell Biology, University of Edinburgh, Edinburgh EH9 3BF, United Kingdom

Abstract

Over the last decade, cross-linking/mass spectrometry (CLMS) has developed into a robust and flexible tool that provides medium-resolution structural information on previously intractable protein complexes and and their interactions. CLMS data describes the proximity of amino acid residues and therefore information on the fold of proteins, and the topology of their complexes. Here, we describe notable successes of this technique and common pipelines. Novel CLMS applications such as *in-cell* cross-linking, probing conformational changes and tertiary structure determination are beginning to make large contributions to molecular biology and the emerging fields of structural systems biology and interactomics.

1. Introduction

Cross-Linking/Mass Spectrometry (CLMS, also abbreviated as CL-MS, XL-MS, CX-MS, or CXMS) has developed in recent years into a robust technology that is accessible to many biochemical laboratories. It is now a standard complementary tool to traditional structural techniques[1] and has benefited from intense methodological development. Pioneering laboratories now focus on expanding CLMS pipelines to generate even greater amounts of information on protein structure, protein complex topology, quantitation of conformational states and analysing protein-protein interactions (PPIs) on a system-wide scale.

The cross-linking reaction adds new covalent bonds between proximal residues. Commonly, this is done with soluble cross-linkers that target surface residues (Table 1). Alternatively, photoactivatable amino acids such as photo-methionine and photo-leucine[2] can be globally incorporated during translation, allowing the protein interior and hydrophobic patches to be probed. Distance restraints are generated by identifying the cross-linked residues and considering the length of the most extended conformation of the cross-linking reagent. These data are then used, often in conjunction with other available structural information, for modelling the conformation of proteins and the topology of their complexes.

The experimental steps in CLMS typically involve the reaction, digestion of cross-linked proteins, enrichment of the cross-linked peptides and tandem mass spectrometry-based data acquisition followed by peptide identification by database searching (Fig. 1). There are many different approaches to each of these steps and therefore many potential ways to combine them into a pipeline[3]. After protein digestion, normally by trypsin, cross-linked peptides are sub-stoichiometric compared to their non-cross-linked linear counterparts as not every protein or peptide is cross-linked in the same way. Detecting cross-linked peptides within these mixtures has been the longstanding bottleneck of the CLMS pipeline.

Improvements in cross-linking reagents, mass spectrometers and database searching algorithms have resulted in the successful application of CLMS to systems with increasing complexity[4–6] and have vastly improved the sensitivity of cross-link detection and/or identification[7]. Major advantages, compared with other structural techniques, are that the cross-linking reaction occurs in solution, it tolerates a large variety of buffer conditions, deals with sample heterogeneity and requires relatively small sample quantities. CLMS played a significant role in some of the most ambitious recent integrated structural biology studies, including the mitochondrial ribosome[8], the mediator-RNA polymerase II complex[9] and the mammoth task of piecing together the membrane-bound nuclear pore complex[10–12].

Proteins and protein complexes are not static entities and CLMS captures the ensemble of their conformations in solution, providing information on dynamics and flexible regions. Quantitative CLMS (QCLMS) developments now allow for direct comparison of cross-links derived from a complex in different states, i.e. in the presence and absence of an effector. Differences in cross-link patterns could be due to conformational differences between the two samples. This contrasts with most other structural techniques, which normally require conformational homogeneity of the sample. CLMS also has the unique potential to contribute to the field of 'structural systems biology' by not only providing high-throughput mapping of cellular PPIs *in vivo* or *in vitro,* but also adding the missing information on protein conformations and interaction topologies to cellular interaction networks.

2. CLMS applications

CLMS applied to protein assemblies

The most established application of CLMS is to investigate the topology of enriched or purified protein complexes (Fig. 2a). Unlike for other structural techniques, the sample does not need to be purified, although significant enrichment of the protein complex will improve cross-link detection, and the quantities required are relatively small (10's-100's µg). The combination of CLMS with other structural and modelling techniques, often termed 'integrative structural biology', has been used to determine the subunit/domain organisation of complexes and assign ambiguous densities in EM maps to specific subunits[13–15]. It helped identify that previous structural work on the CCT chaperonin rings had placed homologous subunits in the wrong order[16,17]. The cross-linking reaction occurs in solution in near-native conditions and so is an ideal complementary technique to address possible artifacts of other experimental approaches such as crystal contacts perturbing protein structure.

Particularly fruitful recent applications of CLMS in combination with high-resolution electron microscopy have provided important structural information on the positions and orientations of subunits for challenging large protein assemblies that had previously eluded structural determination. These structures include: the 26S proteasome holocomplex structure[18], the 40S ribosome-eIF1-eIF3 complex[19], the chromatin-remodelling complex SWR1[20], the INO80 chromatin remodeller[21], the RNA polymerase II-TFIIF complex[13], RNA polymerase II coupled with the pre-mRNA capping complex[22], RNA polymerase II-mediator core initiation complex[9], the mitochondrial ribosome[8], the yeast spliceosome[23] and the nuclear pore complex[11,12].

CLMS paired with integrative protein structural modelling has spurred the development of specialised restraint-driven pipelines where the cross-links put limitations on potential subunit orientations, restricting the energy landscape to drive sampling in a way that will lead to accurate models[24,25]. Further valuable topological information such as stoichiometries and surface-accessible regions can be integrated from other MS-based technologies such as ion mobility MS, native MS[26] and hydrogen−deuterium exchange MS[27]. Lower-resolution structural models of complexes that are not amenable to standard structural techniques can be built by combining biochemical and genetic evidence with CLMS and available atomic-level information, as work on the SAGA transcription coactivator complex[28] and the SEA complex[29] demonstrate.

CLMS cannot normally discriminate between intra-subunit and inter-subunit cross-links in homomultimeric complexes except for the rare case where a cross-link is identified between two peptides that overlap in sequence and

therefore cannot be from the same protein molecule[30]. An elegant solution to this is to mix stable isotope-labelled subunits e.g. ^{15}N-labelled protein with unlabelled ^{14}N protein[31]. Cross-links between subunits are differentiated as linking a ^{15}N labelled peptide to a ^{14}N peptide that has been used to analyse the oligomeric state of guanylyl cyclase-activating protein 2[32] and p53[33].

CLMS can also be used to confirm that complex topologies derived from *in vitro* experiments correspond to those found *in situ*. In a study on the SMC2/SMC4 subcomplex[34], the purified complex was first analysed by CLMS *in vitro.* To validate that the identified *in vitro* topology occurs in the cell, the detected cross-links were used to direct a targeted mass spectrometry approach in order to find the same cross-links in *in situ* cross-linked chromatin.

The approaches described above rely on a combination of lower density cross-links with higher resolution information about individual subunits. However, such information is not always available for all subunits within a complex. Furthermore, the analysed system may undergo conformational changes or display a dynamic behaviour that cannot be deduced from the structures of the individual subunits.

CLMS for tertiary protein structure determination

A truly non-specific cross-linker could provide distance restraints between surface residues of any type. The potential combinations of cross-linked residues obtained with such a cross-linker means that the data would be challenging to analyse. A compromise is to use heterobifunctional reagents such as sulfosuccinimidyl 4,4-azipentanoate (sulfo-SDA). This cross-linker combines a sulfo-NHS ester and a promiscuous UV photoactivatable diazirine group[7]. One arm of this cross-linker is first anchored on a lysine, serine, threonine or tyrosine and the other is free to link any residue after photoactivation, making data acquisition and analysis simpler than for a truly random cross-linker. Data from this type of cross-linker has been termed high-density cross-linking (HD-CLMS)[35,36]. Nevertheless, the increased sample complexity and the increased search space have so far limited the application of this approach to single proteins.

High-density information from CLMS can be leveraged to facilitate generation of structural models *de novo* without any prior knowledge of the protein structure (Fig. 2b). HD-CLMS derived distance restraints are combined with computational approaches that predict protein folds by restraining the proximity of structural domains, loops and residues. Small numbers of cross-links have previously been used to validate predictions made through computational

approaches[37,38], and have proven to be largely complementary to evolutionary conserved contacts[39]. In the first study to show that it is possible to collect the numbers of restraints with sulfo-SDA required to make sensible tertiary structure predictions, we probed the 66 kDa human serum albumin. This study reported 1495 cross-links that were used in a hybrid modelling pipeline to predict the HSA domains that broadly agreed with the known crystal structure[7].

HD-CLMS has now been incorporated into the Critical Assessment of protein Structure Prediction (CASP), a community effort to test the success of algorithms for predicting structures of proteins for which the structure is known but is yet to be publicly released. HD-CLMS debuted at CASP11 where the inclusion of HD-CLMS restraints produced a modest improvement over other molecular modelling approaches but most importantly, it provided a foundation on which to design further integration of this data[40,41].

Additional complementary cross-links such as those with a shorter spacer length to give tighter distance restraints, or cross-links in the core of the protein by photoactivatable amino acids will provide more restraints for proteins that are difficult to model. HD-CLMS may become an invaluable tool for solving novel structures that have to date evaded standard structural biology techniques.

Quantitative CLMS for comparative studies

Generally, CLMS data has been interpreted under the assumption that the investigated protein or protein complexes are static entities. However, a purified protein can be an ensemble of many different states. Thus, CLMS studies often produce high-confidence cross-links that cannot be explained by the available crystal structures used to benchmark them. Some of these discrepancies may represent conformations that are present in solution but are not available in the PDB[42]. Since cross-linking data can represent a mixture of all of the conformations occuring in the system, a careful analysis of long-distance cross-links can be used to separate these alternative conformations. This approach has led to key insights into the GRK5 interaction with the b2-adrenergic receptor[43].

The analysis of conformational changes by cross-linking is eased by designing comparative analyses[44,45] such as using isotope-labelling techniques, which allow direct comparison of the abundances of specific cross-links between different protein states (Fig. 2c). Early applications of this method were to investigate the conformational changes that take place after spontaneous hydrolysis of a thioester bond in the complement protein C3[46] and the impact of phosphorylation on the conformation of an F-type ATPase, by comparing cross-links derived from complexes treated with either BS3-d0 or BS3-d4[47]. The

relative abundances of the cross-links were compared manually. Since then, efforts have been made to automate this approach with XiQ[45], xTract[48] and incorporation into MaxQuant[49]. Other applications of this technology have shown large conformational rearrangements that occur in the proteasome during particle assembly[50], the regulation of enzymes[48,51,52] and the regulation of protein interactions[53,54].

QCLMS has shown the most success in cases where conformational equilibria can be strongly shifted by, for example, effector binding or post-translational modifications. However, even subtle structural changes that result in altered residue proximities, solvent accessibility or steric hindrances have also been observed to affect the formation and yield of cross-links[55]. Great care should be taken in interpreting cross-link changes if factors affecting cross-link reactivity are changed, for example differences in pH, or factors affecting digestion such as differential post-translational modification.

CLMS applied to proteome-wide scales

CLMS offers the possibility to generate distance restraints across the entire proteome that reveal medium resolution structural information of proteins and their interactions (Fig. 2d). This is extremely ambitious due to the complexity of the starting material. Complexity here refers to the total proteins in the sample and all possible combinations of their resulting peptides that need to be considered during data analysis.

Cellular PPIs range from stable 'core' complexes to very transient interactions. Large-scale proteomics studies that have provided the most comprehensive PPI maps by affinity-purification/mass spectrometry[56,57] or co-fractionation[58,59] do not reveal the physical arrangement of proteins within the identified complexes. CLMS can add this missing topological information to the PPI maps and also capture interactions that may be lost during purification. The number of different proteins, the range of abundances and the post-translational modifications in the cell makes the detection of cross-links for all but the most abundant proteins difficult. Nevertheless, rapid technological advancements in CLMS are now allowing production of PPI networks on starting material that only a few years ago would have been impossibly complex.

In general, there are three different approaches to generating starting material for CLMS-based PPI maps; targeted pull-downs, cell lysates and *in situ* cross-linking.

Pull-down studies that enrich the native complex being investigated combined with on-bead cross-linking can provide topological information that can

225 be used to separate direct binders from background[60–62]. In the case of protein
226 phosphatase 2A (PP2A), pull-downs and on-bead cross-linking were used to
227 disentangle a complicated interaction network that consists of many different
228 isoforms[14]. Whereas particularly transient interactions are lost during these pull-
229 down protocols, there have also been promising studies using *in situ* stabilisation
230 of the tagged protein complexes by cross-linking in their native environment prior
231 to the pull-down. These studies have utilised cross-linkers that can penetrate cell
232 membranes and cross-link protein complexes in cells or in organelles. The
233 tagged proteins can be pulled-down along with their interaction network under
234 denaturing conditions, and this approach has been used most notably in the
235 proteasome interactome[63].
236 CLMS studies applied to the whole soluble proteome in cell lysates have
237 reported several hundreds of PPIs in several species, including human cell
238 lines[6,64–66]. The data generated so far has been sparse and the majority of the
239 cross-links are within abundant and well-characterised complexes such as the
240 nucleosome, ribosome and proteasome. While a single confident inter-protein
241 cross-link is enough to confirm an interaction, it is preferable to get a density of
242 cross-link distance restraints that would provide topological information. One
243 possibility to navigate around this complexity is to cross-link cell lysates simplified
244 by biochemical fractionation[67].
245 Several attempts have been made to produce cross-link-based interaction
246 maps *in situ* in cells or organelles without genetic tagging or pull-downs in order
247 to preserve the important contribution of molecular crowding to *in vivo*
248 interactions, and to maintain the most physiologically relevant conditions. These
249 approaches produce the most complex starting material of all and although the
250 cross-linkers used are often chemically enrichable, here too only moderate
251 numbers of cross-links are identified[4,68–71]. Excitingly, even with current
252 superficial insights, cross-linking intact cells has revealed interesting biology
253 including the interaction dynamics of HSP90[72], interactions between bacterial
254 virulence factors and host proteins[73], and it has suggested supercomplexes in
255 the mitochondrial oxidative phosphorylation system[4,71]. Cross-links can be used
256 to check the proposed folds of homology models of protein structures from
257 species that had not been previously structurally investigated[67].
258 Many novel interactions and previously unknown quaternary structures
259 have been identified in these *in situ* studies but often with insufficient numbers of
260 cross-links to enable modelling of the interacting surfaces or docking. Using the
261 NHS-ester cross-linker DSS, it is possible to identify in the order of 50-100 cross-
262 links on a typical purified protein, so for a human proteome cross-linking
263 experiment, a crude conservative estimate is that >200,000 cross-links are there

to be identified within and between the 4000 most abundant proteins. This represents a significant analytical challenge and demonstrates that improvements in the enrichment and analysis of cross-linked peptides are required.

3. CLMS workflows

The development of standardised reagents and workflows have hugely increased the ease-of-use of CLMS, though the abundance of available workflows may cause confusion to newcomers. Below, we discuss the major integrated CLMS workflows that are streamlined pipelines built around specific cross-linker chemistries and search software. There are now numerous different software solutions available for identifying cross-linked peptides[74]. Regardless of the search software used, the standard method to gauge confidence is FDR estimation by a target-decoy approach[30,75,76]. Reporting standards (Box 1) and data visualisation software (Box 2 and Fig. 3) are easing interpretation and transfer of results necessary for the democratisation of this technique.

Importantly, the chemistry of the cross-linker spacer (the part remaining after a cross-link is formed) can be modified to aid data analysis and give confidence to the identified cross-links. Therefore, the most suitable cross-linker in combination with the analysis pipeline should be carefully considered before embarking on a study.

Universal approach

This is the broadest approach that requires no modification to the cross-linker spacer to aid downstream analysis. This is widely used in combination with standard commercial cross-linkers and is particularly useful for cross-linkers that are currently incompatible with modification of their spacer region, such as photo amino acids[2,77]. Isotope-labelling is not required for identification and can consequently be used for quantitative or comparative studies (Fig. 4a).

This strategy takes advantage of the accuracy with which modern mass spectrometers can record fragmentation spectra. The high mass accuracy limits the chance of false positives by randomly matching signals to the database of potential cross-linked peptides. Many software suites, including StavroX[78], Xlink-Identifier[79] and Xcomb[80] generate a database of potential cross-linked pairwise peptide combinations, but as the number of proteins increases, they can require impractical amounts of computational time. Other approaches use an open modification search strategy where each cross-linked spectrum is searched as two peptides, each with a modification of an unknown mass at an unknown

residue/lysine[81]. Commonly used search strategies also combine the modification search with experimental heuristics that enrich for potential cross-linked peptides computationally to save search time before scoring the spectra, including Xi[82], Plink[65], XLSearch[83], Protein Prospector[84], ECL2[85] and Kojak[86].

Xi, a search software of our design, first linearises spectra *in silico*[82] and then performs a simple linear search on that spectra, which outputs a list of candidates for one of the linked peptides. The linearisation step uses an empirical understanding of the fragmentation spectra of cross-linked peptides, revealing cross-linked fragments due to their charge and fractional mass. Some peptides do not fragment well so we do not necessarily expect to identify both cross-linked peptides in this first step. For each of the candidates, the mass of this first peptide and the mass of the cross-linker are known. Consequently, a mass of the second peptide can be calculated (precursor mass – (cross-linker + peptide 1)). Every peptide that fits this mass is considered as a candidate peptide and all predicted pairs are used to score the whole spectrum. Xi identifies up to ~20% more links compared to algorithms that only consider Lys-Lys linkages by also searching side reactions of the amine-reactive cross-linker with serine, threonine and tyrosine residues[87]. Xi also permits the search for products of photo-cross-linking to obtain high-density CLMS data[7]. A false discovery rate (FDR) for this data can be calculated using a target-decoy approach with our software xiFDR[75] that has an integrated option to further increase identifications through noise reduction.

Labelled cross-linker approach

This approach is designed to indicate in the first mass spectrometry stage (MS1) which peptides contain a cross-linker, whether those are cross-linked peptides or simply cross-linker modified peptides (i.e. mono-links), and to reveal cross-linked fragments (Fig. 4b). This helps to identify cross-links, which is especially beneficial for confident cross-link identification in cases where high-accuracy mass spectrometers are not available. It also simplifies the data analysis workflow. Several labelling strategies exist[88–90]. The most common is to use a 1:1 mixture of heavy and light isotope-labelled cross-linkers during the reaction. This produces a characteristic doublet in MS1 spectra that is recognised by search software. Several search software take advantage of this approach including: Hekate[91], StavroX[78] and the popular xQuest[92].

For example, xQuest identifies the linked peptides using the fragmentation spectra (MS2) of both the light- and heavy-labelled MS1 precursors. Linear peptide fragments, which do not contain the labelled cross-linker are common

peaks in both fragmentation spectra and can be used to map the two cross-linked peptides[93] (Fig. 4b). This reduces the potential search space by generating lists of candidate peptides that can then be scored against the full MS2 spectra. xProphet, a software package that is paired with xQuest, then calculates FDRs using a target-decoy approach[76].

However ,this elegant approach has disadvantages. Peptides modified with the isotopically heavy and normal cross-linker can differ in retention time on the reverse-phase liquid chromatography in line with the mass spectrometer, which can make it difficult to identify the pairs. Requiring both the heavy and light precursors to be picked for fragmentation can cause issues in complex samples. Labelling with a 1:1 ratio of light/heavy cross-linker causes the intensity of the cross-linked peptides to be halved relative to the unlabelled linear peptides in the sample. This decreases the likelihood that their precursors are selected for fragmentation Labelling also increases the complexity of the MS1 spectral space, which can negatively affect identification rates. Moreover, use of labelled cross-linkers to identify linked peptides prevents their use of this for QCLMS (see 'Quantitative CLMS for comparative studies').

MS2-cleavable cross-linker approach

Cross-linked peptides are large and branched, which gives rise to complicated fragmentation spectra and uneven fragmentation. The large number of potential peptide combinations in conjunction with the often poor fragmentation of one of the cross-linked peptides can make it difficult to confidently identify the two peptides. Separating the two peptides in the mass spectrometer simplifies the analysis to identifying two individual linear peptides (Fig. 4c).

Consequently, considerable effort has been expended on creating cross-linkers that are cleavable in the mass spectrometer so that spectra simply correspond to two modified peptides[94–105] .

MS2-cleavable cross-linker approaches tend to cleave asymmetrically during mass spectrometric fragmentation, which provides distinctive cross-link specific product ions that report the presence of cross-linked peptides. DSSO (see Table 1), for example, cleaves asymmetrically leaving a distinctive group of 4 peaks after CID (collision-induced dissociation) fragmentation: peptide A with the longer arm of the cleaved DSSO, peptide B with the shorter arm and *vice versa*[94] (Fig. 4c). These can be selected for further fragmentation (MS3), which simplifies the spectra to that of a simple modified peptide with a measured parent m/z and therefore aids identification. The use of MS3 requires an instrument that is capable of doing MS3 - and this technique takes longer duty cycles than the

MS2-only approaches discussed above. Acquisition approaches for these cross-linkers have subsequently been designed by several laboratories along with their respective search software such as ICC-CLASS[106], MeroX[107], X-links/Blinks[108,109] and XlinkX2.0[6,66].

In the XlinkX pipeline, a mixed approach is recommended by performing sequential CID and ETD (electron-transfer dissociation) fragmentation on each precursor. Fragment ions displaying the characteristic doublets of DSSO are selected for further MS3 analysis[66]. The information from all three of these fragmentation approaches is then combined for identification. This data integration circumvents a large drawback with this approach whereby the MS2-cleavage of the cross-linker is often inefficient and therefore doublet peaks are not always obvious for selection for MS3.

4. Future prospects

The field of CLMS is providing powerful tools to molecular biologists to aid structural biology and interactome research. CLMS has matured into a core technique for *in vitro* structural studies that is capable of delivering medium resolution information to complement classical atomic resolution structural biology techniques and computational modelling. CLMS results from a purified protein complex can be generated and analysed in under a week by most proteomics core facilities. The addition of QCLMS, which may soon become routine, means that structural differences caused by conformational changes or mutations can be assessed in solution. HD-CLMS has demonstrated potential for aiding the characterisation of tertiary protein structure in combination with computational modelling.

In situ studies of PPIs and protein structures is the next phase of the CLMS revolution. As work continues towards acquiring data of greater depth, CLMS may eventually become a widely-used quantitative *in cell* structural technique to monitor interactions and conformational changes simultaneously. Several hundreds to thousands of cross-linked residue pairs can be identified from purified protein complexes so the few thousands of cross-linked peptides that have so far been detected in ‘proteome-wide’ studies are only the tip of the iceberg. Improved cross-linked peptide enrichment, cross-linker chemistries, and further progress in data acquisition and analysis will allow us to map many tens of thousands of cross-links within the cell and elucidate protein topologies *in situ* at a true proteome-wide scale. One envisions a time in the near future when CLMS will be used to routinely map entire protein interactomes and their dynamics during biological processes such as cellular differentiation,

development and the transition from health to disease.

Acknowledgements

We apologise to our colleagues for the many missing citations to their work which were omitted solely to keep this introductory review lean or due to our own ignorance. We would like to thank the reviewers for their very helpful suggestions.

Funding info

This work was supported by the Einstein Foundation, the DFG [RA 2365/4-1], and the Wellcome Trust through a Senior Research Fellowship to JR [103139]. The Wellcome Centre for Cell Biology is supported by core funding from the Wellcome Trust [203149].

Cross-linker	Spacer length	Residue reactivity
BS3/DSS	11.4 Å	K/S/T/Y/N-terminus - K/S/T/Y
DSSO	10.3 Å	K/S/T/Y/N-terminus - K/S/T/Y
EDC	0 Å	D/E - K/S/T/Y
DMTMM + spacer dihydrazine	Depends on spacer	D/E - D/E
SDA	3.4 Å	K/S/T/Y/N-terminus - Any

Table 1| A selection of commonly used soluble cross-linkers and their reactivities. *Cross-linkers comprise two elements: reactive end-groups and a spacer. The reactive groups dictate which residues are targeted and thereby the amount of spatial information that can be obtained. Cross-linkers that target specific reactive groups generate fewer potential combinations of cross-linked residues, which limits the amount of structural information but also eases the data analysis. The spacer region of the cross-linker is what remains when two residues are cross-linked and largely determines the workflow to be used including the MS acquisition method and data analysis. The spacer also*

influences the spatial resolution and data density. A longer spacer will allow more residue combinations to fall within a cross-linker's range, which may be beneficial if the main goal is to identify proteins (not necessarily the residues) that are in close proximity. However, a longer spacer also reduces the information value of a cross-link as a cross-link only produces an upper bound distance restraint.

Box 1: Reporting standards

The field has not yet settled on minimal reporting standards. This has disadvantages for assessing publications and reusing data for modelling. The proteomics standards initiative (HUPO-PSI) has developed an XML-based reporting standard for proteomics data, mzIdentML (http://www.psidev.info/mzidentml), which from version 1.2 includes CLMS[110]. Published mass spectrometric raw data should be deposited in ProteomeXchange. When reporting results, there is a need for disambiguation with regards to the term 'cross-link', which is often used without clear distinction for peptide spectral matches (PSMs), peptide pairs and residue pairs. This is confusing when trying to assess data quantity, as PSMs and peptide pairs contain redundant information, and can lead to serious flaws in data reliability[75]. As a minimum, authors must define the term and use it consistently.

Box 2: Data visualisation and interpretation

Commonly, laboratories interested in applying CLMS to their structural problem will collaborate with laboratories that specialise in CLMS or proteomics core facilities to generate data. To facilitate the accessibility and interpretation of CLMS results for experts and non-experts, software has been developed by several laboratories for visualising the identified cross-links and the mass spectra that led to their identification[91,111–117]. Cross-linking studies provide many levels of information; residue-residue links, 3D structural information (when integrated with atomic level information) and protein-protein interactions, which has required bespoke visualisations (Fig. 3).

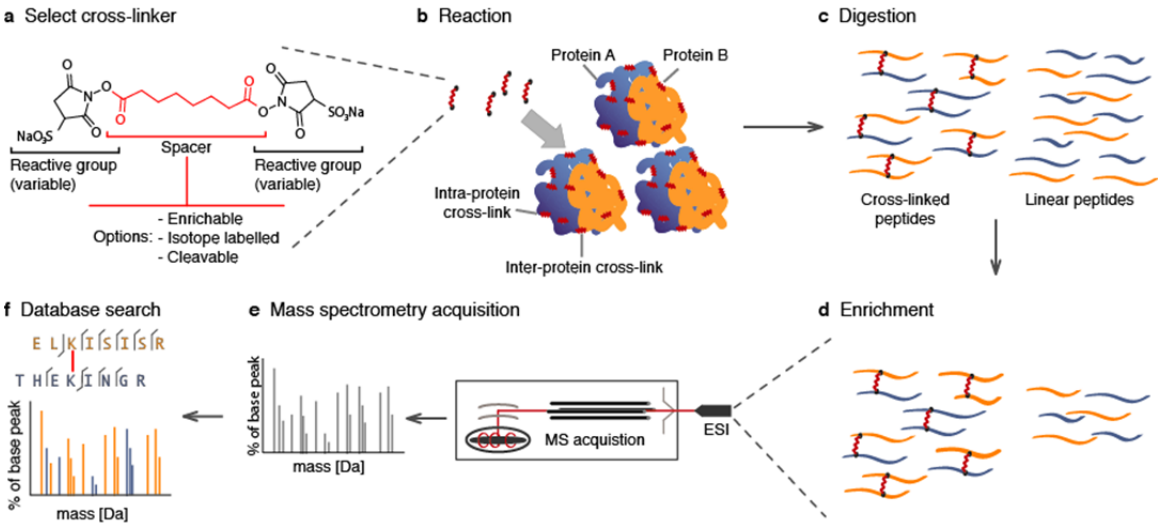


Fig. 1| General CLMS workflow. a, *Cross-linkers can contain various chemistries and lengths. Depending on the experimental workflow used (see CLMS data analysis and integrated workflows), the cross-linker spacer may need to be cleavable, labelled or have enrichable moieties.* Reactive groups are also variable (Table 1). **b,** *Concentrations and reaction times need to be empirically checked 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.* **d,** *After digestion cross-linked peptides are often enriched due to the large abundance of linear peptides for all applications more complex than a single protein. This can be achieved by chromatographic methods such as size exclusion chromatography or strong cation exchange chromatography.* **e,** *MS/MS pipelines have been designed to increase the likelihood of selecting cross-linked peptide precursors for fragmentation.* **f,** *Various search software solutions are described in the main text that can identify the two linked peptides from the spectra.*

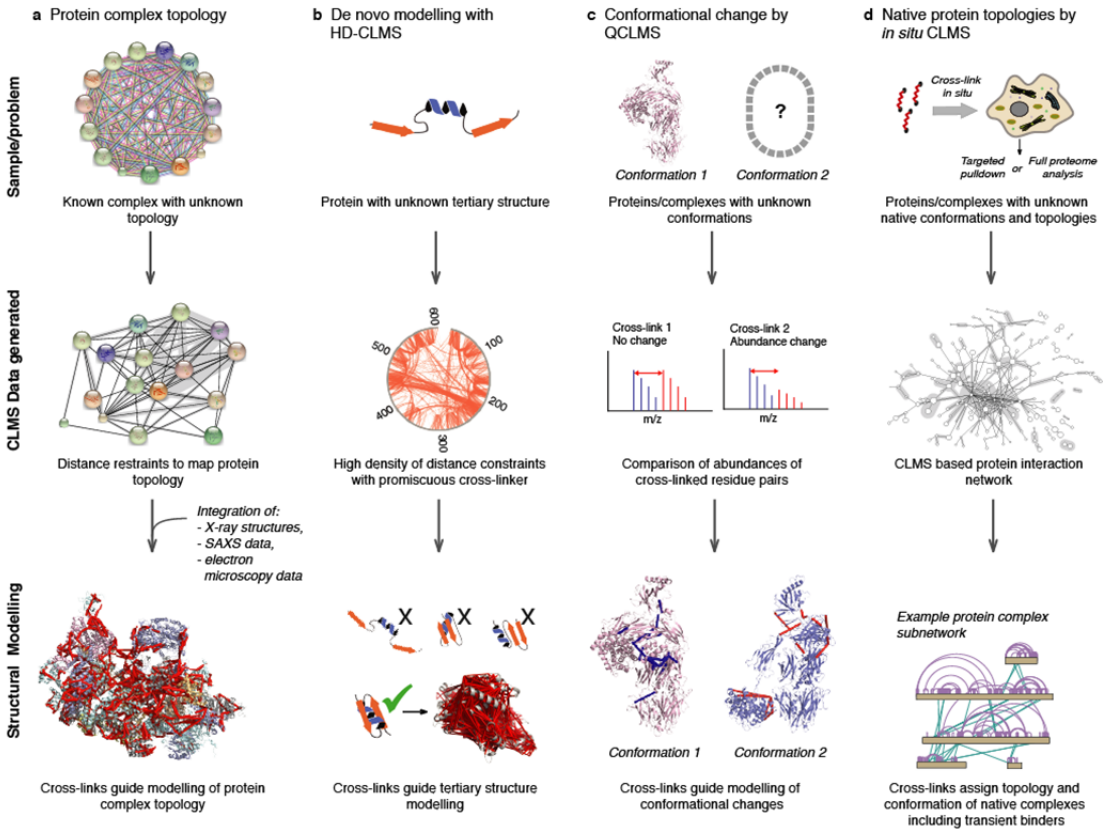


Fig. 2| CLMS applications. a, *Distance restraints from CLMS allows docking of subunits of protein complexes. Together with other structural techniques such as cryo-EM can be used to provide medium to high resolution structural information for previously intractable complexes.* **b**, *High-density cross-linking using photo-activatable cross-linkers, while limited by sample complexity provides data density that can be used to guide algorithms that fold the tertiary structure of proteins using de novo or database aided approaches.* **c**, *QCLMS can describe structural differences between two protein/protein complex conformations by using isotope labelling to compare the abundances of cross-links detected in different samples.* **d**, *The complexity of samples analysed by CLMS has increased in recent years to include cell lysates and in situ analyses of organelles and whole cells. It is also possible to target specific interaction networks by cross-linking the cells and then pull-down a specific protein and its interactors.*

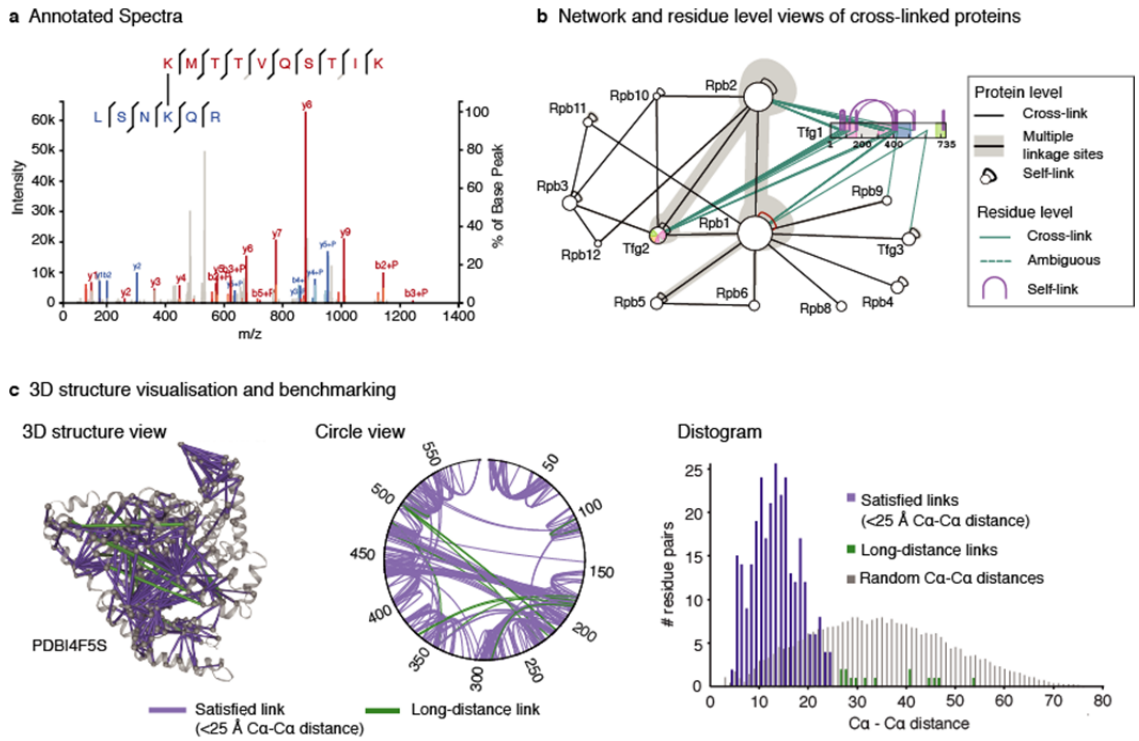


Fig. 3| Visualisation solutions for CLMS data. a, *Spectra identified as cross-linked peptides can be manually assessed. To aid this, spectral viewer XiSPEC[118] highlights fragment peaks have been used to identify the linked peptides and their fragmentation coverage.* **b,** *Cross-linked proteins can be visualised using node and edge graphs to display interconnectivity of proteins. Intralinks and interlinks are visualised at residue and domain resolution by representing the proteins as bars or concatenated into circle plots, as used in XiNet[111] and XVis[112]. Cross-linked edges can also be colored to aid interpretation of qCLMS data.* **c,** *Mapping of cross-links on known 3D structures or homology models can score and validate cross-links and show those that violate the distance restraints, as implemented in xiNET[111], Xlink Analyzer[113] , XLink-DB[114], Xwalk[115], CLMSVault[116], ProXL[117] and Hekate[91]. This same information can be visualised as a distogram - a histogram that shows the distribution of cross-link distances in the data[111]. Normally a distance restraint is considered satisfied if the euclidean distance between Cα's is less than the cross-linker spacer, plus the side chains, plus an empirically derived short distance accounting for flexibility in the peptide backbone*

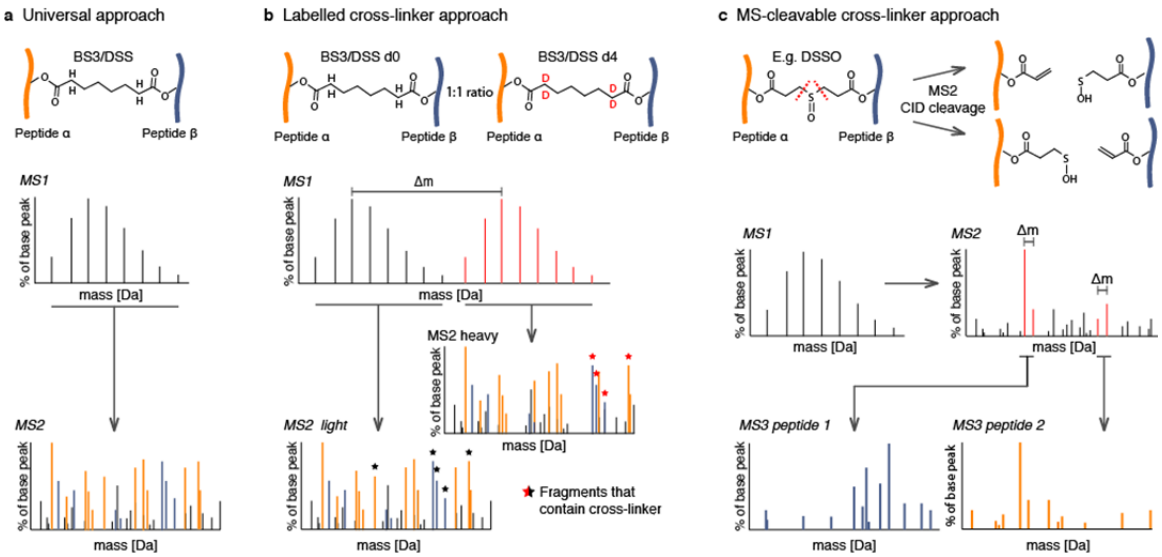


Fig. 4| CLMS data acquisition and analysis workflows. a, *The 'universal approach' uses cross-linkers with simple spacers. The fragmentation spectra are a mixture of fragments from both peptides and must be resolved during database search.* **b,** *The sample is cross-linked with a mixture of cross-linker and its heavy isotope labeled counterpart. Here usually all MS1 precursors are fragmented and during database searching the MS1 doublets (peaks separated by the mass difference between the cross-linkers) are used to indicate spectra that contain cross-linker. When the MS2 spectra from heavy and light labeled peptides are compared, peaks with a shifted mass (here indicated with stars) indicate fragments that contain the cross-linker and can therefore simplify data analysis.* **c,** *The 'MS cleavable cross-linker approach' can be done in one of two ways. The cross-linker is cleavable in the MS2, often asymmetrically, so it produces fragments that indicate the separate masses of the two cross-linked peptides. These MS2 spectra can be used alone for the database search or, more commonly, the characteristic doublet ions from the asymmetrically cleaved cross-linker are selected for MS3 fragmentation of each peptide separately during acquisition. These MS3 spectra are then used for database searching along with the deduced masses of the parent peptides.*

References

1. Leitner, A., Faini, M., Stengel, F. & Aebersold, R. Crosslinking and Mass Spectrometry: An Integrated Technology to Understand the Structure and Function of Molecular Machines. *Trends Biochem. Sci.* **41,** 20–32 (2016).
2. Pham, N. D., Parker, R. B. & Kohler, J. J. Photocrosslinking approaches to interactome mapping. *Curr. Opin. Chem. Biol.* **17,** 90–101 (2013).
3. Yu, C. & Huang, L. Cross-Linking Mass Spectrometry: An Emerging Technology for Interactomics and Structural Biology. *Anal. Chem.* **90,** 144–165 (2018).
4. Schweppe, D. K. *et al.* Mitochondrial protein interactome elucidated by chemical cross-linking mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* (2017). doi:10.1073/pnas.1617220114
5. Chavez, J. D. *et al.* Quantitative interactome analysis reveals a chemoresistant edgotype. *Nat. Commun.* **6,** 7928 (2015).
6. Liu, F., Rijkers, D. T. S., Post, H. & Heck, A. J. R. Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry. *Nat. Methods* **12,** 1179–1184 (2015).
7. Belsom, A., Schneider, M., Fischer, L., Brock, O. & Rappsilber, J. Serum Albumin Domain Structures in Human Blood Serum by Mass Spectrometry and Computational Biology. *Mol. Cell. Proteomics* **15,** 1105–1116 (2016).
8. Greber, B. J. *et al.* Ribosome. The complete structure of the 55S mammalian mitochondrial ribosome. *Science* **348,** 303–308 (2015).
9. Plaschka, C. *et al.* Architecture of the RNA polymerase II-Mediator core initiation complex. *Nature* **518,** 376–380 (2015).
10. Bui, K. H. *et al.* Integrated structural analysis of the human nuclear pore complex scaffold. *Cell* **155,** 1233–1243 (2013).
11. Kosinski, J. *et al.* Molecular architecture of the inner ring scaffold of the human nuclear pore complex. *Science* **352,** 363–365 (2016).
12. Kim, S. J. *et al.* Integrative structure and functional anatomy of a nuclear pore complex. *Nature* (2018). doi:10.1038/nature26003
13. Chen, Z. A. *et al.* Architecture of the RNA polymerase II–TFIIF complex revealed by cross□linking and mass spectrometry. *EMBO J.* **29,** 717–726 (2010).
14. Herzog, F. *et al.* Structural probing of a protein phosphatase 2A network by chemical cross-linking and mass spectrometry. *Science* **337,** 1348–1352 (2012).
15. Joachimiak, L. A., Walzthoeni, T., Liu, C. W., Aebersold, R. & Frydman, J. The structural basis of substrate recognition by the eukaryotic chaperonin TRiC/CCT. *Cell* **159,** 1042–1055 (2014).
16. Kalisman, N., Adams, C. M. & Levitt, M. Subunit order of eukaryotic TRiC/CCT chaperonin by cross-linking, mass spectrometry, and combinatorial homology modeling. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 2884–2889 (2012).
17. Leitner, A. *et al.* The molecular architecture of the eukaryotic chaperonin TRiC/CCT. *Structure* **20,** 814–825 (2012).

18. Lasker, K. *et al.* Molecular architecture of the 26S proteasome holocomplex determined by an integrative approach. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 1380–1387 (2012).
19. Erzberger, J. P. *et al.* Molecular architecture of the 40S□eIF1□eIF3 translation initiation complex. *Cell* **158,** 1123–1135 (2014).
20. Nguyen, V. Q. *et al.* Molecular architecture of the ATP-dependent chromatin-remodeling complex SWR1. *Cell* **154,** 1220–1231 (2013).
21. Tosi, A. *et al.* Structure and subunit topology of the INO80 chromatin remodeler and its nucleosome complex. *Cell* **154,** 1207–1219 (2013).
22. Martinez-Rucobo, F. W. *et al.* Molecular Basis of Transcription-Coupled Pre-mRNA Capping. *Mol. Cell* **58,** 1079–1089 (2015).
23. Yan, C. *et al.* Structure of a yeast spliceosome at 3.6-angstrom resolution. *Science* **349,** 1182–1191 (2015).
24. Ferber, M. *et al.* Automated structure modeling of large protein assemblies using crosslinks as distance restraints. *Nat. Methods* **13,** 515–520 (2016).
25. Karaca, E., Rodrigues, J. P. G. L. M., Graziadei, A., Bonvin, A. M. J. J. & Carlomagno, T. M3: an integrative framework for structure determination of molecular machines. *Nat. Methods* **14,** 897–902 (2017).
26. Politis, A. *et al.* A mass spectrometry-based hybrid method for structural modeling of protein complexes. *Nat. Methods* **11,** 403–406 (2014).
27. Konermann, L., Pan, J. & Liu, Y.-H. Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chem. Soc. Rev.* **40,** 1224–1234 (2011).
28. Han, Y., Luo, J., Ranish, J. & Hahn, S. Architecture of the Saccharomyces cerevisiae SAGA transcription coactivator complex. *EMBO J.* **33,** 2534–2546 (2014).
29. Algret, R. *et al.* Molecular architecture and function of the SEA complex, a modulator of the TORC1 pathway. *Mol. Cell. Proteomics* **13,** 2855–2870 (2014).
30. Maiolica, A. *et al.* Structural analysis of multiprotein complexes by cross-linking, mass spectrometry, and database searching. *Mol. Cell. Proteomics* **6,** 2200–2211 (2007).
31. Merkley, E. D. *et al.* Mixed-Isotope Labeling with LC-IMS-MS for Characterization of Protein–Protein Interactions by Chemical Cross-Linking. *J. Am. Soc. Mass Spectrom.* **24,** 444–449 (2013).
32. Pettelkau, J. *et al.* Structural analysis of guanylyl cyclase-activating protein-2 (GCAP-2) homodimer by stable isotope-labeling, chemical cross-linking, and mass spectrometry. *J. Am. Soc. Mass Spectrom.* **24,** 1969–1979 (2013).
33. Arlt, C., Ihling, C. H. & Sinz, A. Structure of full-length p53 tumor suppressor probed by chemical cross-linking and mass spectrometry. *Proteomics* **15,** 2746–2755 (2015).
34. Barysz, H. *et al.* Three-dimensional topology of the SMC2/SMC4 subcomplex from chicken condensin I revealed by cross-linking and molecular modelling. *Open Biol.* **5,** 150005 (2015).

35. Belsom, A., Schneider, M., Brock, O. & Rappsilber, J. Blind Evaluation of Hybrid Protein Structure Analysis Methods based on Cross-Linking. *Trends Biochem. Sci.* **41,** 564–567 (2016).
36. Schneider, M., Belsom, A. & Rappsilber, J. Protein Tertiary Structure by Crosslinking/Mass Spectrometry. *Trends Biochem. Sci.* **43,** 157–169 (2018).
37. Young, M. M. *et al.* High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 5802–5806 (2000).
38. Singh, P., Nakatani, E., Goodlett, D. R. & Catalano, C. E. A pseudo-atomic model for the capsid shell of bacteriophage lambda using chemical cross-linking/mass spectrometry and molecular modeling. *J. Mol. Biol.* **425,** 3378–3388 (2013).
39. Dos Santos, R. N. *et al.* Enhancing protein fold determination by exploring the complementary information of chemical cross-linking and coevolutionary signals. *Bioinformatics* (2018). doi:10.1093/bioinformatics/bty074
40. Schneider, M., Belsom, A., Rappsilber, J. & Brock, O. Blind testing of cross-linking/mass spectrometry hybrid methods in CASP11. *Proteins* **84 Suppl 1,** 152–163 (2016).
41. Belsom, A. *et al.* Blind testing cross-linking/mass spectrometry under the auspices of the 11th critical assessment of methods of protein structure prediction (CASP11). *Wellcome Open Res* **1,** 24 (2016).
42. Ding, Y.-H. *et al.* Modeling Protein Excited-state Structures from 'Over-length' Chemical Cross-links. *J. Biol. Chem.* **292,** 1187–1196 (2017).
43. Komolov, K. E. *et al.* Structural and Functional Analysis of a β2-Adrenergic Receptor Complex with GRK5. *Cell* **169,** 407–421.e16 (2017).
44. Chen, Z. A. & Rappsilber, J. Protein dynamics in solution by quantitative cross-linking/mass spectrometry. *Trends Biochem. Sci.* **IN PRESS,** (2018).
45. Fischer, L., Chen, Z. A. & Rappsilber, J. Quantitative cross-linking/mass spectrometry using isotope-labelled cross-linkers. *J. Proteomics* **88,** 120–128 (2013).
46. Chen, Z. A. *et al.* Structure of Complement C3(H2O) Revealed By Quantitative Cross-Linking/Mass Spectrometry And Modeling. *Mol. Cell. Proteomics* **15,** 2730–2743 (2016).
47. Schmidt, C. *et al.* Comparative cross-linking and mass spectrometry of an intact F-type ATPase suggest a role for phosphorylation. *Nat. Commun.* **4,** 1985 (2013).
48. Walzthoeni, T. *et al.* xTract: software for characterizing conformational changes of protein complexes by quantitative cross-linking mass spectrometry. *Nat. Methods* **12,** 1185–1190 (2015).
49. Chen, Z. A., Fischer, L., Cox, J. & Rappsilber, J. Quantitative Cross-linking/Mass Spectrometry Using Isotope-labeled Cross-linkers and MaxQuant. *Mol. Cell. Proteomics* **15,** 2769–2778 (2016).
50. Tomko, R. J., Jr *et al.* A Single α Helix Drives Extensive Remodeling of the Proteasome Lid and Completion of Regulatory Particle Assembly. *Cell* **163,** 432–

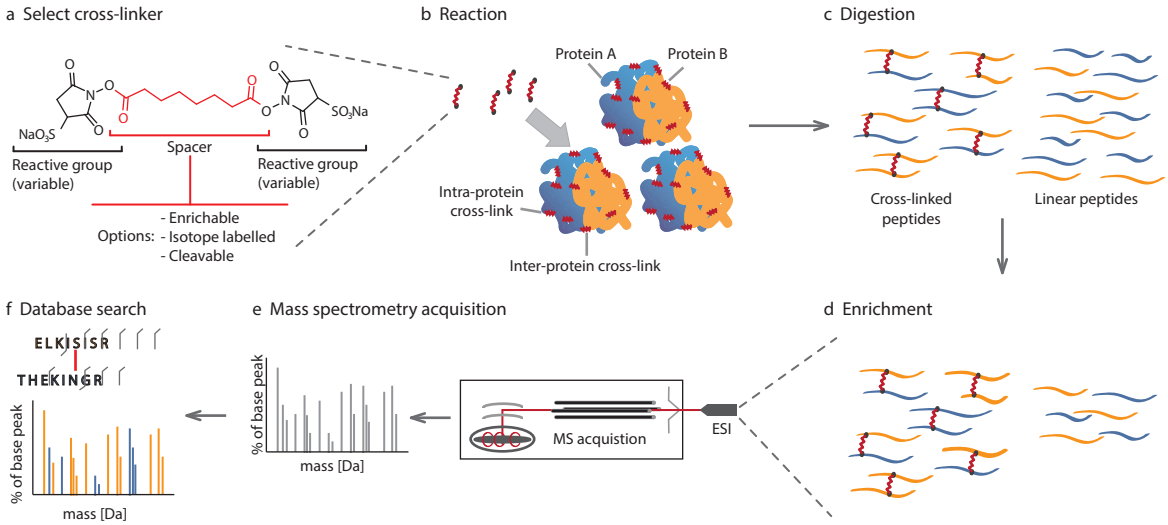
670 444 (2015).
671 51. Yu, C. *et al.* Gln40 deamidation blocks structural reconfiguration and activation of
672 SCF ubiquitin ligase complex by Nedd8. *Nat. Commun.* **6,** 10053 (2015).
673 52. Boelt, S. G. *et al.* Mapping the Ca(2+) induced structural change in calreticulin. *J.*
674 *Proteomics* **142,** 138–148 (2016).
675 53. Beilsten-Edmands, V. *et al.* eIF2 interactions with initiator tRNA and eIF2B are
676 regulated by post-translational modifications and conformational dynamics. *Cell*
677 *Discov* **1,** 15020 (2015).
678 54. Koehler, C. *et al.* Genetic code expansion for multiprotein complex engineering. *Nat.*
679 *Methods* **13,** 997–1000 (2016).
680 55. Chen, Z. *et al.* Quantitative cross-linking/mass spectrometry reveals subtle protein
681 conformational changes. *Wellcome Open Res* **1,** 5 (2016).
682 56. Gavin, A.-C. *et al.* Proteome survey reveals modularity of the yeast cell machinery.
683 *Nature* **440,** 631–636 (2006).
684 57. Krogan, N. J. *et al.* Global landscape of protein complexes in the yeast
685 Saccharomyces cerevisiae. *Nature* **440,** 637–643 (2006).
686 58. Havugimana, P. C. *et al.* A census of human soluble protein complexes. *Cell* **150,**
687 1068–1081 (2012).
688 59. Kristensen, A. R., Gsponer, J. & Foster, L. J. A high-throughput approach for
689 measuring temporal changes in the interactome. *Nat. Methods* **9,** 907–909 (2012).
690 60. Makowski, M. M., Willems, E., Jansen, P. W. T. C. & Vermeulen, M. Cross-linking
691 immunoprecipitation-MS (xIP-MS): Topological Analysis of Chromatin-associated
692 Protein Complexes Using Single Affinity Purification. *Mol. Cell. Proteomics* **15,** 854–
693 865 (2016).
694 61. Shi, Y. *et al.* A strategy for dissecting the architectures of native macromolecular
695 assemblies. *Nat. Methods* **12,** 1135–1138 (2015).
696 62. Häupl, B., Ihling, C. H. & Sinz, A. Protein Interaction Network of Human Protein
697 Kinase D2 Revealed by Chemical Cross-Linking/Mass Spectrometry. *J. Proteome*
698 *Res.* **15,** 3686–3699 (2016).
699 63. Wang, X. *et al.* Molecular Details Underlying Dynamic Structures and Regulation of
700 the Human 26S Proteasome. *Mol. Cell. Proteomics* **16,** 840–854 (2017).
701 64. Tan, D. *et al.* Trifunctional cross-linker for mapping protein-protein interaction
702 networks and comparing protein conformational states. *Elife* **5,** (2016).
703 65. Yang, B. *et al.* Identification of cross-linked peptides from complex samples. *Nat.*
704 *Methods* **9,** 904–906 (2012).
705 66. Liu, F., Lössl, P., Scheltema, R., Viner, R. & Heck, A. J. R. Optimized fragmentation
706 schemes and data analysis strategies for proteome-wide cross-link identification.
707 *Nat. Commun.* **8,** 15473 (2017).
708 67. Kastritis, P. L. *et al.* Capturing protein communities by structural proteomics in a
709 thermophilic eukaryote. *Mol. Syst. Biol.* **13,** 936 (2017).
710 68. Kaake, R. M. *et al.* A New in Vivo Cross-linking Mass Spectrometry Platform to
711 Define Protein–Protein Interactions in Living Cells. *Mol. Cell. Proteomics* **13,** 3533–

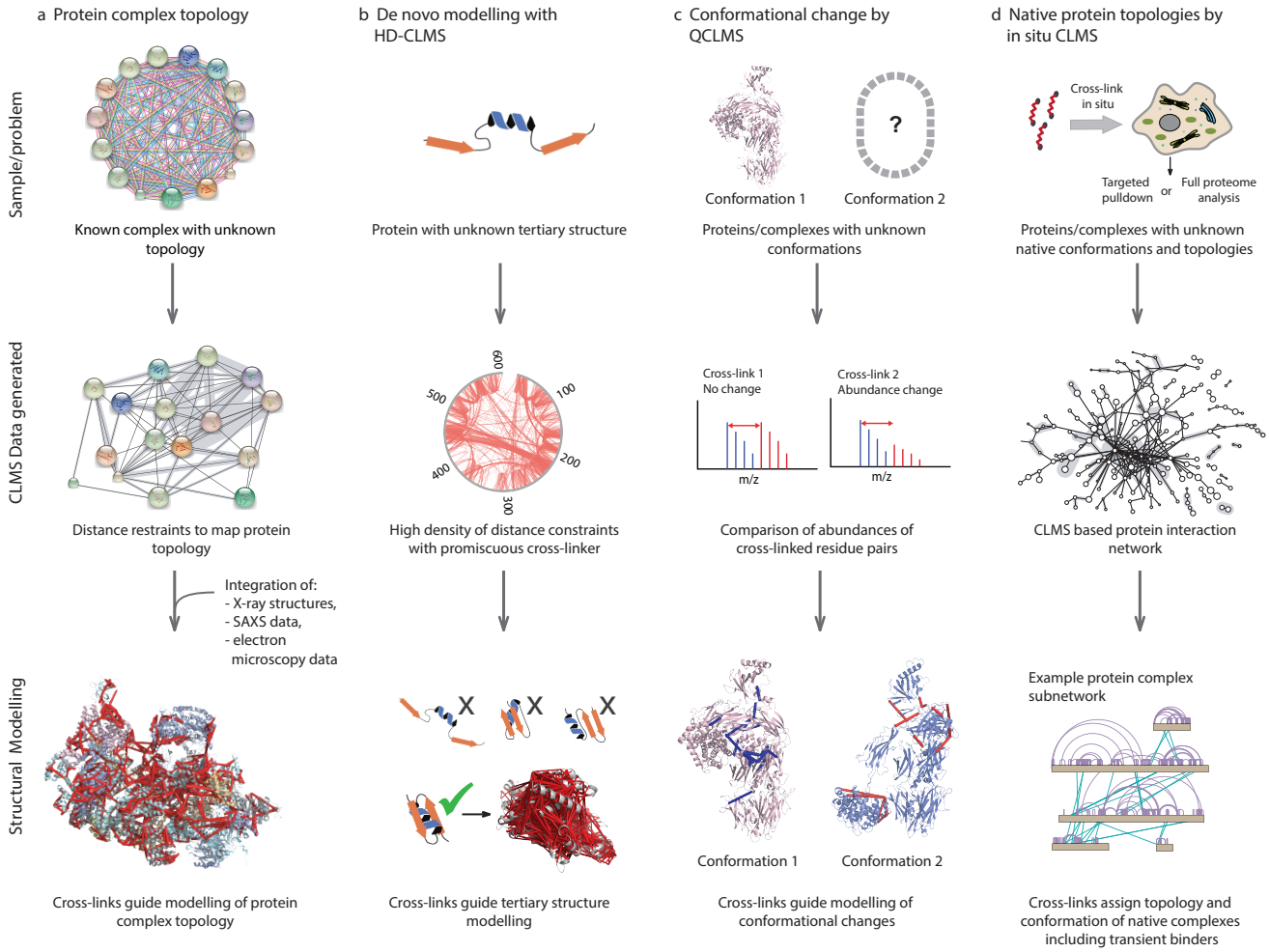
3543 (2014).
69. Chavez, J. D., Weisbrod, C. R., Zheng, C., Eng, J. K. & Bruce, J. E. Protein interactions, post-translational modifications and topologies in human cells. *Mol. Cell. Proteomics* **12,** 1451–1467 (2013).
70. Wu, X. *et al.* In vivo protein interaction network analysis reveals porin-localized antibiotic inactivation in Acinetobacter baumannii strain AB5075. *Nat. Commun.* **7,** 13414 (2016).
71. Liu, F., Lössl, P., Rabbitts, B. M., Balaban, R. S. & Heck, A. J. R. The interactome of intact mitochondria by cross-linking mass spectrometry provides evidence for coexisting respiratory supercomplexes. *Mol. Cell. Proteomics* **17,** 216–232 (2018).
72. Chavez, J. D., Schweppe, D. K., Eng, J. K. & Bruce, J. E. In Vivo Conformational Dynamics of Hsp90 and Its Interactors. *Cell Chem Biol* **23,** 716–726 (2016).
73. Schweppe, D. K. *et al.* Host-Microbe Protein Interactions during Bacterial Infection. *Chem. Biol.* **22,** 1521–1530 (2015).
74. Tran, B. Q., Goodlett, D. R. & Goo, Y. A. Advances in protein complex analysis by chemical cross-linking coupled with mass spectrometry (CXMS) and bioinformatics. *Biochim. Biophys. Acta* **1864,** 123–129 (2016).
75. Fischer, L. & Rappsilber, J. On the Quirks of Error Estimation in Cross-Linking/ Mass Spectrometry. *Anal. Chem.* (2017). doi:10.1021/acs.analchem.6b03745
76. Walzthoeni, T. *et al.* False discovery rate estimation for cross-linked peptides identified by mass spectrometry. *Nat. Methods* **9,** 901–903 (2012).
77. Suchanek, M., Radzikowska, A. & Thiele, C. Photo-leucine and photo-methionine allow identification of protein-protein interactions in living cells. *Nat. Methods* **2,** 261–267 (2005).
78. Götze, M. *et al.* StavroX—A Software for Analyzing Crosslinked Products in Protein Interaction Studies. *J. Am. Soc. Mass Spectrom.* **23,** 76–87 (2012).
79. Du, X. *et al.* Xlink-identifier: an automated data analysis platform for confident identifications of chemically cross-linked peptides using tandem mass spectrometry. *J. Proteome Res.* **10,** 923–931 (2011).
80. Panchaud, A., Singh, P., Shaffer, S. A. & Goodlett, D. R. xComb: a cross-linked peptide database approach to protein-protein interaction analysis. *J. Proteome Res.* **9,** 2508–2515 (2010).
81. Singh, P. *et al.* Characterization of protein cross-links via mass spectrometry and an open-modification search strategy. *Anal. Chem.* **80,** 8799–8806 (2008).
82. Giese, S. H., Fischer, L. & Rappsilber, J. A Study into the Collision-induced Dissociation (CID) Behavior of Cross-Linked Peptides. *Mol. Cell. Proteomics* **15,** 1094–1104 (2016).
83. Ji, C., Li, S., Reilly, J. P., Radivojac, P. & Tang, H. XLSearch: a Probabilistic Database Search Algorithm for Identifying Cross-Linked Peptides. *J. Proteome Res.* **15,** 1830–1841 (2016).
84. Trnka, M. J., Baker, P. R., Robinson, P. J. J., Burlingame, A. L. & Chalkley, R. J. Matching cross-linked peptide spectra: only as good as the worse identification. *Mol.*

Cell. Proteomics **13,** 420–434 (2014).
85. Yu, F., Li, N. & Yu, W. Exhaustively Identifying Cross-Linked Peptides with a Linear Computational Complexity. *J. Proteome Res.* **16,** 3942–3952 (2017).
86. Hoopmann, M. R. *et al.* Kojak: efficient analysis of chemically cross-linked protein complexes. *J. Proteome Res.* **14,** 2190–2198 (2015).
87. Yuan, Z. *et al.* Structural basis of Mcm2-7 replicative helicase loading by ORC-Cdc6 and Cdt1. *Nat. Struct. Mol. Biol.* **24,** 316–324 (2017).
88. Petrotchenko, E. V., Olkhovik, V. K. & Borchers, C. H. Isotopically coded cleavable cross-linker for studying protein-protein interaction and protein complexes. *Mol. Cell. Proteomics* **4,** 1167–1179 (2005).
89. Zelter, A. *et al.* Isotope signatures allow identification of chemically cross-linked peptides by mass spectrometry: a novel method to determine interresidue distances in protein structures through cross-linking. *J. Proteome Res.* **9,** 3583–3589 (2010).
90. Ihling, C. *et al.* Isotope-labeled cross-linkers and Fourier transform ion cyclotron resonance mass spectrometry for structural analysis of a protein/peptide complex. *J. Am. Soc. Mass Spectrom.* **17,** 1100–1113 (2006).
91. Holding, A. N., Lamers, M. H., Stephens, E. & Skehel, J. M. Hekate: software suite for the mass spectrometric analysis and three-dimensional visualization of cross-linked protein samples. *J. Proteome Res.* **12,** 5923–5933 (2013).
92. Rinner, O. *et al.* Identification of cross-linked peptides from large sequence databases. *Nat. Methods* **5,** 315–318 (2008).
93. 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 the xQuest/xProphet software pipeline. *Nat. Protoc.* **9,** 120–137 (2013).
94. Kao, A. *et al.* Development of a novel cross-linking strategy for fast and accurate identification of cross-linked peptides of protein complexes. *Mol. Cell. Proteomics* **10,** (2011).
95. Soderblom, E. J. & Goshe, M. B. Collision-induced dissociative chemical cross-linking reagents and methodology: Applications to protein structural characterization using tandem mass spectrometry analysis. *Anal. Chem.* **78,** 8059–8068 (2006).
96. Buncherd, H., Roseboom, W., de Koning, L. J., de Koster, C. G. & de Jong, L. A gas phase cleavage reaction of cross-linked peptides for protein complex topology studies by peptide fragment fingerprinting from large sequence database. *J. Proteomics* **108,** 65–77 (2014).
97. Müller, M. Q., Dreiocker, F., Ihling, C. H., Schäfer, M. & Sinz, A. Cleavable cross-linker for protein structure analysis: reliable identification of cross-linking products by tandem MS. *Anal. Chem.* **82,** 6958–6968 (2010).
98. Yu, C., Kandur, W., Kao, A., Rychnovsky, S. & Huang, L. Developing new isotope-coded mass spectrometry-cleavable cross-linkers for elucidating protein structures. *Anal. Chem.* **86,** 2099–2106 (2014).
99. Dreiocker, F., Müller, M. Q., Sinz, A. & Schäfer, M. Collision-induced dissociative chemical cross-linking reagent for protein structure characterization: applied Edman

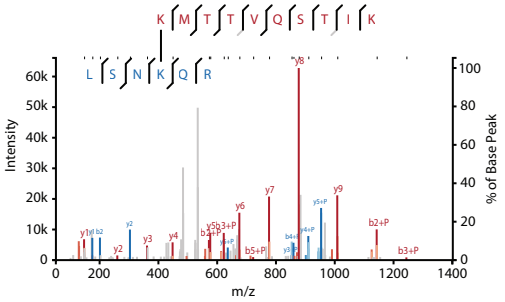
chemistry in the gas phase. *J. Mass Spectrom.* **45,** 178–189 (2010).
100. Petrotchenko, E. V., Serpa, J. J. & Borchers, C. H. An isotopically coded CID-cleavable biotinylated cross-linker for structural proteomics. *Mol. Cell. Proteomics* **10,** (2011).
101. Hage, C., Falvo, F., Schäfer, M. & Sinz, A. Novel Concepts of MS-Cleavable Cross-linkers for Improved Peptide Structure Analysis. *J. Am. Soc. Mass Spectrom.* (2017).
102. Hage, C., Iacobucci, C., Rehkamp, A., Arlt, C. & Sinz, A. The First Zero-Length Mass Spectrometry-Cleavable Cross-Linker for Protein Structure Analysis. *Angew. Chem. Int. Ed Engl.* (2017).
103. Clifford-Nunn, B., Showalter, H. D. H. & Andrews, P. C. Quaternary diamines as mass spectrometry cleavable crosslinkers for protein interactions. *J. Am. Soc. Mass Spectrom.* **23,** 201–212 (2012).
104. Chakrabarty, J. K., Naik, A. G., Fessler, M. B., Munske, G. R. & Chowdhury, S. M. Differential Tandem Mass Spectrometry-Based Cross-Linker: A New Approach for High Confidence in Identifying Protein Cross-Linking. *Anal. Chem.* **88,** 10215–10222 (2016).
105. Tang, X. & Bruce, J. E. A new cross-linking strategy: protein interaction reporter (PIR) technology for protein–protein interaction studies. *Mol. Biosyst.* **6,** 939–947 (2010).
106. Petrotchenko, E. V. & Borchers, C. H. ICC-CLASS: isotopically-coded cleavable crosslinking analysis software suite. *BMC Bioinformatics* **11,** 64 (2010).
107. Götze, M. *et al.* Automated assignment of MS/MS cleavable cross-links in protein 3D-structure analysis. *J. Am. Soc. Mass Spectrom.* **26,** 83–97 (2015).
108. Hoopmann, M. R., Weisbrod, C. R. & Bruce, J. E. Improved strategies for rapid identification of chemically cross-linked peptides using protein interaction reporter technology. *J. Proteome Res.* **9,** 6323–6333 (2010).
109. Anderson, G. A., Tolic, N., Tang, X., Zheng, C. & Bruce, J. E. Informatics strategies for large-scale novel cross-linking analysis. *J. Proteome Res.* **6,** 3412–3421 (2007).
110. Vizcaíno, J. A. *et al.* The mzIdentML Data Standard Version 1.2, Supporting Advances in Proteome Informatics. *Mol. Cell. Proteomics* **16,** 1275–1285 (2017).
111. Combe, C. W., Fischer, L. & Rappsilber, J. xiNET: cross-link network maps with residue resolution. *Mol. Cell. Proteomics* **14,** 1137–1147 (2015).
112. 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.* **43,** W362–9 (2015).
113. Kosinski, J. *et al.* Xlink Analyzer: software for analysis and visualization of cross-linking data in the context of three-dimensional structures. *J. Struct. Biol.* **189,** 177–183 (2015).
114. Zheng, C. *et al.* XLink-DB: Database and Software Tools for Storing and Visualizing Protein Interaction Topology Data. *J. Proteome Res.* **12,** 1989–1995 (2013).
115. Kahraman, A., Malmström, L. & Aebersold, R. Xwalk: computing and visualizing

distances in cross-linking experiments. *Bioinformatics* **27,** 2163–2164 (2011).
116. Courcelles, M. *et al.* CLMSVault: A Software Suite for Protein Cross-Linking Mass-Spectrometry Data Analysis and Visualization. *J. Proteome Res.* (2017).
117. Riffle, M., Jaschob, D., Zelter, A. & Davis, T. N. 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 (2016).
118. Kolbowski, L., Combe, C. & Rappsilber, J. xiSPEC: web-based visualization, analysis and sharing of proteomics data. *Nucleic Acids Res.* (2018).

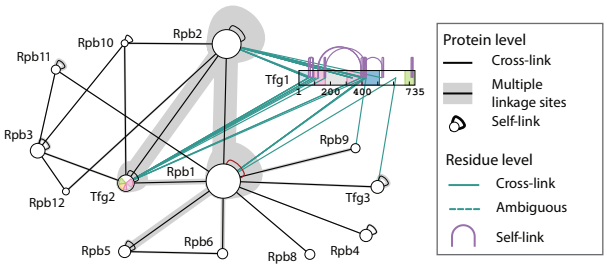

a Select cross-linker
NaO3S
SO3Na
Spacer
Reactive group (variable)
Reactive group (variable)
- Enrichable
Options: - Isotope labelled
- Cleavable
b Reaction
Protein A
Protein B
Intra-protein cross-link
Inter-protein cross-link
c Digestion
Cross-linked peptides
Linear peptides
d Enrichment
ESI
e Mass spectrometry acquisition
MS acquistion
% of base peak
mass [Da]
f Database search
ELKISIS R
THEKINGR
% of base peak
mass [Da]


a Protein complex topology
b De novo modelling with HD-CLMS
c Conformational change by QCLMS
d Native protein topologies by in situ CLMS
Sample/problem
Known complex with unknown topology
Protein with unknown tertiary structure
Conformation 1
Conformation 2
Proteins/complexes with unknown conformations
Cross-link in situ
Targeted pulldown
or
Full proteome analysis
Proteins/complexes with unknown native conformations and topologies
CLMS Data generated
Distance restraints to map protein topology
600
100
200
300
400
500
High density of distance constraints with promiscuous cross-linker
Cross-link 1
No change
Cross-link 2
Abundance change
m/z
m/z
Comparison of abundances of cross-linked residue pairs
CLMS based protein interaction network
Integration of:
- X-ray structures,
- SAXS data,
- electron microscopy data
Structural Modelling
Cross-links guide modelling of protein complex topology
Cross-links guide tertiary structure modelling
Conformation 1
Conformation 2
Cross-links guide modelling of conformational changes
Example protein complex subnetwork
Cross-links assign topology and conformation of native complexes including transient binders

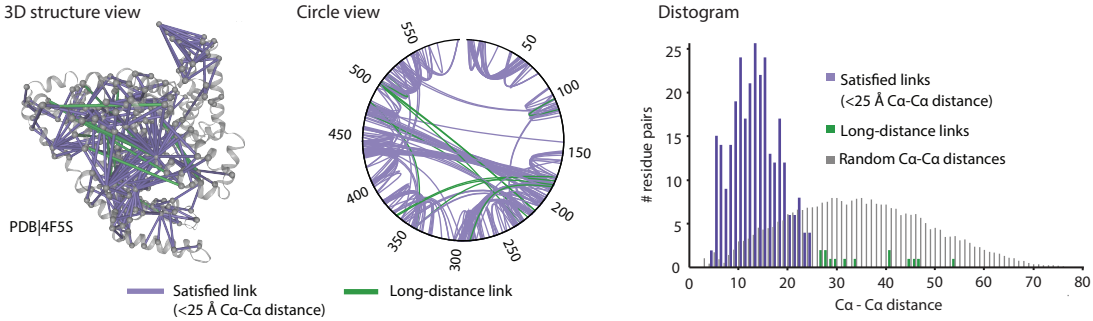
a Annotated Spectra

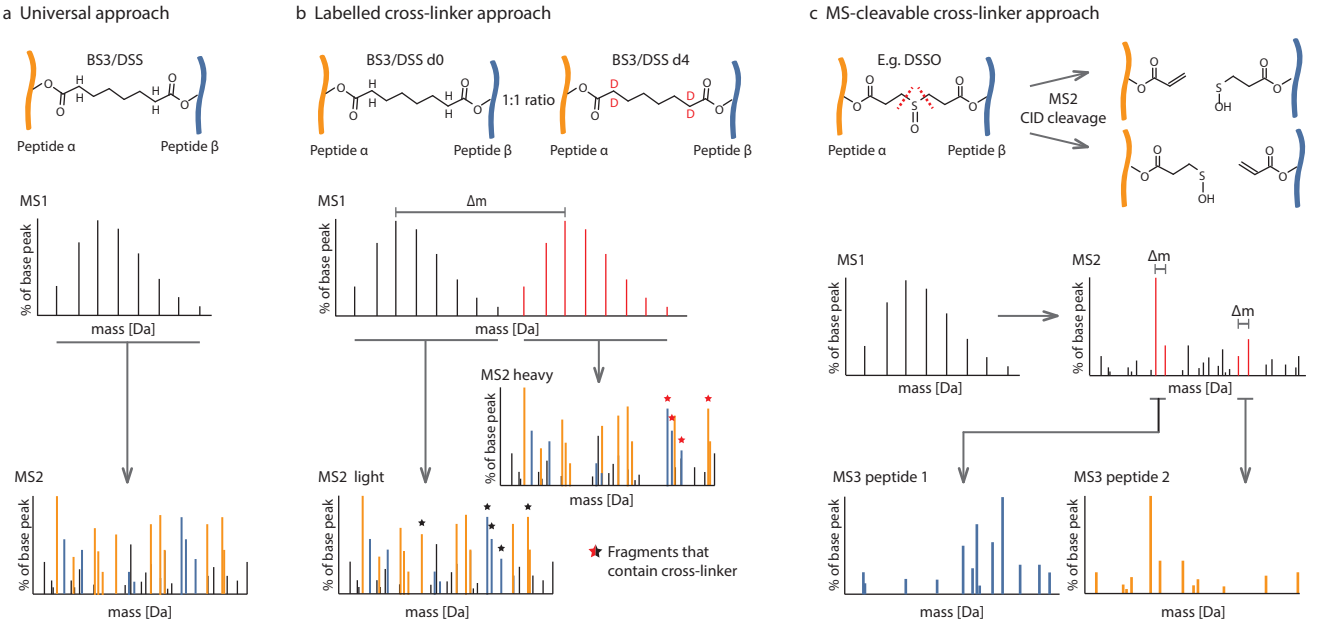


b Network and residue level views of cross-linked proteins



c 3D structure visualisation and benchmarking




a Universal approach
BS3/DSS
Peptide α
Peptide β
MS1
% of base peak
mass [Da]
MS2
% of base peak
mass [Da]
b Labelled cross-linker approach
BS3/DSS d0
1:1 ratio
BS3/DSS d4
Peptide α
Peptide β
Peptide α
Peptide β
MS1
Δm
% of base peak
mass [Da]
MS2 heavy
% of base peak
mass [Da]
MS2 light
% of base peak
mass [Da]
Fragments that contain cross-linker
c MS-cleavable cross-linker approach
E.g. DSSO
Peptide α
Peptide β
MS2
CID cleavage
MS1
% of base peak
mass [Da]
MS2
Δm
Δm
% of base peak
mass [Da]
MS3 peptide 1
% of base peak
mass [Da]
MS3 peptide 2
% of base peak
mass [Da]