Leveraging crosslinking mass spectrometry in structural and cell biology

Leveraging crosslinking mass spectrometry in structural and cell biology

Andrea Graziadei[1], Juri Rappsilber [1,2]

[1] Bioanalytics, Institute of Biotechnology, Technische Universität Berlin, 13355 Berlin, Germany
[2] Wellcome Centre for Cell Biology, University of Edinburgh, Max Born Crescent, Edinburgh, EH9 3BF, UK

Abstract
Crosslinking mass spectrometry (crosslinking-MS) is a flexible tool that is capable of providing structural insights into protein conformation and protein-protein interactions (PPIs). Its medium-resolution residue-residue distance restraints have been used to validate protein structures proposed by other methods, and also helped derive models of protein complexes by integrative structural biology approaches. The use of crosslinking-MS in integrative approaches is underpinned by progress in estimating error rates in crosslinking-MS data and in combining these data with other structural information. The flexible and high-throughput nature of crosslinking-MS has allowed it to complement the ongoing resolution revolution in electron microscopy by providing system-wide residue-residue distance restraints, especially where high-resolution cannot be achieved, such as in flexible regions or systems. Here, we review how crosslinking-MS information has been leveraged in structural model validation and integrative modeling. Crosslinking-MS has also been a key technology for cell biology studies and structural systems biology where, in conjunction with cryo-electron tomography, it can provide structural and mechanistic insights directly *in situ*.

Introduction

Our understanding of life at the molecular level rests on a thorough structural and mechanistic characterisation of its macromolecular components and their interactions. Much of this understanding has been achieved through the main structural biology techniques— X-ray crystallography, nuclear magnetic resonance and electron microscopy (EM). The high resolution models obtained by these methods have allowed us to reconstruct several of the key mechanisms of cellular life in atomic detail. These methods, which require pure or near-pure samples, have relied on the development of heterologous protein expression and advanced biochemical reconstitution methods, as well as innovations in automation and data analysis, to obtain models of molecular machines of increasing complexity.

In recent years, the determination of structures of complex macromolecular assemblies has also been tackled by integrative approaches (Alber et al., 2007; Koukos and Bonvin, 2020; Ward et al., 2013), where traditional structural techniques are complemented by a wealth of information from other sources, such as homology modeling, mutagenesis, biochemical interaction analyses, scattering experiments, surface labeling techniques, and crosslinking mass spectrometry (crosslinking-MS).

Crosslinking-MS is a method that adds a chemical crosslinker to proteins to introduce a covalent bond between nearby residues. The proteins are subsequently digested and the linked peptides are identified by tandem liquid chromatography-mass spectrometry (LC-MS) to reveal which residues were in proximity in the original three-dimensional structure. Thus, the identification of crosslinked residue pairs provides spatial restraint information. Crosslinking-MS has grown over the past decade into a versatile approach with applications ranging from structural biology to cell biology. The residue-level pairwise distance restraints identified in crosslinking-MS studies have been successfully combined with high-resolution structural information from traditional structural biology techniques to obtain models of large protein complexes (Agafonov et al., 2016; von Appen et al., 2015; Chen et al., 2010; Dauden et al., 2017; Erzberger et al., 2014; Gutierrez et al., 2020; Kim et al., 2018; Lasker et al., 2012; Xie et al., 2021; Yan et al., 2019). These crosslinking-MS restraints provide information to identify proteins in electron microscopy densities, confirm residue assignment in modelled regions, localise flexible regions that cannot be resolved, and confirm structures in non-purified and recombinant systems (Casañal et al., 2019; Schmidt and Urlaub, 2017).

Exciting developments in workflow and data analysis have led to the use of crosslinking-MS as a tool in discovery-based studies aiming to identify and characterise the interfaces in systems as complex as cell lysates (Liu et al., 2015), isolated cellular components (Linden et al., 2020; Wittig et al., 2021), whole cells (Chavez et al., 2016; O'Reilly et al., 2020), and tissues (Chavez et al., 2018). Importantly, these studies do not require any genetic modification of the system to introduce protein tags or labels. The high sensitivity of modern mass spectrometers, combined with developments in sample preparation techniques and a solid theoretical framework for handling ambiguity, error, and confidence in crosslinking-MS data has unlocked the potential of crosslinking-MS as a tool in the discovery and structural characterisation of PPIs at the system level at a high throughput (Lenz et al., 2021).

The "resolution revolution" of cryo-electron microscopy and tomography and the corresponding improvements in particle reconstruction and classification methods has catapulted electron microscopy into the method of choice for solving large macromolecular assemblies in atomic and near-atomic detail (Fernandez-Leiro and Scheres, 2016;

Kühlbrandt, 2014; Schur, 2019). This new EM-based possibility to obtain high- and medium-resolution three-dimensional reconstructions from highly complex samples, or even from whole cells, has brought forward a new era in structural biology. Whenever high-resolution information is wholly or partially missing, crosslinking-MS can provide critical distance restraints identifying and placing biomolecules within electron microscopy maps. Moreover, it can characterise flexible regions where density is missing altogether, and identify protein-protein interactions, including from unexpected proteins. In this context, crosslinking-MS has thrived as a tool for model selection, model generation and characterisation of novel interactions in integrative structural biology approaches. This has required the development of a conceptual framework for handling error and confidence in the crosslinking-MS data itself, but also of ways to visualize crosslinking networks and crosslinks on structures. Finally, the use of crosslinks in model scoring, or as restraints to drive the sampling and model generation phase, is underpinned by the development of integrative structural biology software, and of scoring functions capable of integrating physical restraints with the information and uncertainty derived from multiple experimental methods.

The potential of this approach for an *in situ* structural characterisation of the cell at high resolution (in-cell structural biology) is already apparent. Progress towards tackling the huge range of protein and biomolecule abundance, complexity and flexibility found in cells will bring even more diverse applications of this technology.

Crosslinking-MS experimental design: matching chemistry, system and strategy

The objective of crosslinking-MS is to provide distance restraints covering all relevant areas of a protein or complex and to capture the spatial and dynamic information on the system in solution. In a typical crosslinking-MS experiment, a soluble crosslinker is added to a sample ranging in complexity from a purified protein to a whole tissue (O'Reilly and Rappsilber, 2018; Sinz, 2018; Tang et al., 2021; Yu and Huang, 2018). By reacting with specific chemical moieties in the sample, crosslinkers introduce new covalent bonds, which report on the proximity of the chemical groups being probed. Crosslinked samples are then digested by a protease (typically trypsin), and the resulting crosslinked peptides are analysed by mass spectrometry. Specialised search software is then used to identify the two crosslinked peptides from the usually chimeric fragmentation patterns.

Designing a crosslinking-MS experiment involves choosing the appropriate workflow for studying the system of interest and the research question at hand. Crosslinking-MS workflows have been optimised extensively, and care should be taken in selecting the appropriate crosslinker, digestion, enrichment, acquisition and data analysis strategy for a given sample (Fig.1). In particular, employing crosslinking-MS for structural biology of a purified system may lead to different choices at several steps of the workflow than when tackling large-scale protein-protein interaction studies (Chavez et al., 2019; Götze et al., 2019; Iacobucci et al., 2018; Kaake et al., 2014; Liu et al., 2015; Mendes et al., 2019). If the goal of the study is to compare abundances of residue-residue contacts or PPIs under different conditions, quantitation of the precursor and spectral intensities is required. Quantitative crosslinking-MS may then be used to characterise relative abundances of conformations or interactions (Chavez et al., 2015; Chen et al., 2016; Huang and Kim, 2009; Kukacka et al., 2015; Walzthoeni et al., 2015; Zheng et al., 2016).

For the scope of this review, we will focus on crosslinking-MS approaches to identify crosslinks between peptides, though it is worth noting that the identification of peptide-nucleic acid crosslinks, which are often employed in genomic and transcriptomic experiments, is an area of active development (Dorn et al., 2017; Kramer et al., 2014).

First, we consider choosing a crosslinker. Crosslinkers comprise two distinct parts (Belsom and Rappsilber, 2021; Chavez and Bruce, 2019; Yu and Huang, 2018). The first is the reactive groups, each with their own reaction specificities, half-lifes, and kinetics, that specify which residues can be linked (Table 1). Their other important feature is the spacer region— the part that remains after the residues are crosslinked. The length of the spacer dictates the maximum distance that may be probed in the crosslinking MS experiment.

From a structural analysis perspective, reactive group chemistry is a crucial parameter to consider (Belsom and Rappsilber, 2021; Ziemianowicz et al., 2019). NHS-ester based crosslinkers, with their long half-life but relatively high specificity (Anjaneyulu and Staros, 1987; Mädler et al., 2009), provide a good signal-to-noise. For this reason, they have proven very popular in the study of purified complexes (Piersimoni and Sinz, 2020), large assemblies and whole systems, despite the fact that relatively few crosslinks are obtained per protein. Their narrow specificity leads to manageable database sizes in the computational search for crosslink-spectrum matching (see below), which in turn allows the user to tackle systems comprising hundreds of proteins. On the other hand, short-lived, unspecific crosslinkers based on diazirine chemistry lead to increased chemical complexity of the crosslinked peptides and lower abundance of each crosslink. They increase data density (Belsom et al., 2016) but reduce signal-to-noise and increase the dimensions of the computational search space, making their application in systems beyond purified protein complexes challenging.

New families of crosslinkers under development use diverse reactive groups, leading to a wide range of half-lives, specificities and reactivities. As an example, the sulfur fluoride exchange (SuFex) reagents have a long half-life and a low reactivity (Yang et al., 2018).

Crosslinker spacers have been developed to span multiple distances, from the “zero-length” reaction provided by EDC, over the ~11 Å alkyl spacer used for protein interface mapping with DSS/BS3, to the 60 Å of the protein interaction reporter (PIR) (Tang and Bruce, 2010; Wu et al., 2016), used in probing large-scale interactions in cells. The length of the spacer then influences what residue-residue distances may be probed, and can be modelled in various ways when it comes to data interpretation (see below). The spacer may also incorporate additional features that facilitate the analysis, such as affinity tags (Steigenberger et al., 2019; Trester-Zedlitz et al., 2003), or groups allowing for cleavage in the gas phase (Kao et al., 2011; Müller et al., 2010; Yu et al., 2020), which aid mass spectrometric identification. DSSO and DSBU are popular improvements of traditional NHS-ester crosslinkers with spacers that can be cleaved in the gas phase. Gas phase cleavage generates stubs that give rise to characteristic duplet signals (Matzinger and Mechtler, 2021; Steigenberger et al., 2020). Because these fragments can improve the identification of the crosslinked peptides, DSSO and DSBU are often used when the protein mixture is complex. MS-cleavable diazirine crosslinkers have also been recently developed (Gutierrez et al., 2021).

Another possible feature of the crosslinker is isotope labelling, which is particularly useful in quantitative or comparative experiments. In such workflows, samples under one condition are labelled with heavy isotope crosslinker (e.g. ^{2}H or ^{13}C) and mixed 1:1 with samples from another condition carrying the light crosslinker (Fischer et al., 2013; Schmidt and Robinson, 2014). Alternatively, isotope labelling may be performed on the proteins themselves (Chavez et al., 2015), or using isobaric labeling reagents (Yu et al., 2016). The ratio of crosslinked peptide pair signals at MS1, MS2 or MS3 level may then be used to derive abundance ratios of conformers or interactions.

In order to not introduce artefacts, crosslinking is performed in substoichiometric amounts, so that each protein molecule may only contain a single or a few crosslinked peptide

pairs. A crosslinker titration is therefore necessary to optimise the conditions that maximise crosslinking while not producing higher-order aggregation or artefacts.

Digestion strategies have also been carefully optimised to increase sequence coverage: traditional trypsin digestion has been complemented with LysC (Gonzalez-Lozano et al., 2020), elastase (Dau et al., 2019) or additional proteases (Leitner et al., 2012; Mendes et al., 2019).

Because of the low abundance of crosslinked peptide pairs relative to linear and crosslinker-modified linear peptides, their detection in complex samples requires chromatographic enrichment prior to LC-MS acquisition. Various groups have successfully used a single chromatographic enrichment step on purified systems crosslinked with NHS-ester based crosslinkers. Crosslinked peptides have been separated from linear peptides by a variety of chromatographic techniques, including size exclusion, strong cation exchange and hydrophilic strong anion exchange chromatography (Chen et al., 2010; Fritzsche et al., 2012; Leitner et al., 2012; O'Reilly et al., 2020; Ritorto et al., 2013). Certain crosslinkers, such as PIR (Tang and Bruce, 2010), have tags in their spacer region, allowing for affinity-based enrichment of crosslinked peptide pairs and crosslinker-modified peptides. For organelle-wide or whole-cell crosslinking studies, multidimensional fractionation has been employed (e.g. (O'Reilly et al., 2020)). This results in the further enrichment of crosslinked peptide pairs, and in their separation across another chromatographic dimension, thereby reducing sample complexity and increasing identification rates, albeit at the cost of large increases in measurement time and sample amount requirements.

Finally, the crosslinking-MS community has worked on several strategies for obtaining and selecting the highest number of crosslinked peptide pair species during LC-MS acquisition, such as optimised fragmentation parameters (Kolbowski et al., 2017; Liu et al., 2017a). In particular, as crosslinked peptides tend to have higher charge states than their linear counterparts (Maiolica et al., 2007; Rinner et al., 2008), charge state selection has proven a highly successful strategy to limit the acquisition of linear peptides (Chen et al., 2010). Recently, the development of Field Asymmetric Ion Mobility Spectrometry (FAIMS) enabled researchers to perform multidimensional separation (by reversed phase and by size) directly at the spectrometer itself (Schnirch et al., 2020).

Error estimation in crosslinking-MS

In order to use crosslinking-MS information to build interpretable models of protein complexes or protein-protein interaction networks, the error associated with the identification of residue pairs and PPIs has to be accurately quantified and reported. Moreover, a known error rate allows for the integration of crosslinking-MS information with other sources of data (Schneidman-Duhovny et al., 2014).

The identification of crosslinked residue pairs from MS fragmentation spectra relies on database search, i.e. on a computational search of the peaks in the spectra against a database of all proteins that may be present in the sample. Each spectrum may be matched to a linear peptide coming from a single protein, in a peptide-spectrum match (PSM), or to a crosslinked peptide pair, in a crosslink-spectrum match (CSM). There are several popular crosslink search software programs available (Chen et al., 2019; Götze et al., 2019; Hoopmann et al., 2015; Leitner et al., 2014; Mendes et al., 2019), which use different scoring functions to assign a goodness-of-fit between spectrum and peptide sequence to each PSM and CSM. For a CSM this requires the assignment of a match quality to both peptides and also to the combined pair. Guidelines have been developed for avoiding spectral misassignment (Iacobucci and Sinz, 2017).

To convert goodness-of-fit (or score) into confidence requires knowledge of the false positive rate (Fischer and Rappsilber, 2017; Walzthoeni et al., 2012). In order to derive an overall false-discovery rate (FDR) for the detected CSMs, crosslink search engines rely on the target-decoy approach (Maiolica et al., 2007; Walzthoeni et al., 2012), in which the database is supplemented by a decoy database of protein sequences that are not present in the sample. The score distribution of matches to this decoy database should model the score distribution of random matches (CSMs) to the target database. Typically, the decoy database is made up of the original sequences reversed, or reversed and swapped, to model the original or target proteins in search space size and amino acid composition. Thus, each spectrum may be matched by a linear peptide, a decoy linear peptide, a true target-target peptide pair, a target-decoy, or a decoy-decoy peptide pair. Note that co-eluting linear peptides can also associate non-covalently (not chemically linked together) and may thus mimic a crosslink (Giese et al., 2019).

The number of decoys at any particular confidence cutoff may be then used to derive an FDR at the CSM level. The search space, and therefore the chance of random matching, is fundamentally different between “self” crosslinks (crosslinks within a protein sequence) and “heteromeric” crosslinks (crosslinks between different proteins). Thus, self and heteromeric crosslinks must be handled separately in FDR calculations (Lenz et al., 2021). Indeed, a high ratio of heteromeric to self links in a dataset is indicative of higher error rate than the stated FDR (Keller et al., 2019).

Because of the stochastic nature of fragmentation and signal-to-noise in a spectrometer, the same peptide pair may be supported by multiple CSMs, each with its own confidence (Fischer and Rappsilber, 2017). Therefore, the aggregation of CSMs into crosslinked peptide pairs, and of peptide pairs into residue pairs (the same residue pair may come from peptides with different proteolytic cleavage patterns) requires appropriate error propagation techniques (Fig.2A). The same principle applies in crosslinking-MS-based PPI studies, where residue-residue pairs are aggregated into PPIs in order to derive a network with a known false discovery rate (de Jong et al., 2021; Lenz et al., 2021).

It is important to know at which level the FDR has been estimated (CSM, peptide pair, residue pair or PPI), as this will have large effects on their usefulness in structural analysis. Typically, protein modeling studies employ residue-residue pairs as restraints, and so FDR control should be applied at this level (Fischer and Rappsilber, 2017). Using FDR estimation at lower levels will significantly underestimate the identification error of residue pairs and PPIs (Fig. 2A). In a recent study, a 4% error threshold at CSM level led to an actual 17% error at the residue pair level, and a 35% error in the resulting PPIs (Lenz et al., 2021). Note that estimating crosslink identification error based on distance distribution on known structures may also lead to severe underestimation of the error (Yugandhar et al., 2020). This is because the likelihood of a random match occurring within a known interface in a large sequence database is low: therefore, the rate of misidentified crosslinks with both ends in a known structure is much lower than that of links having only one end or none at all in a region of known structure. Furthermore, the error in crosslink identification is not uniformly distributed. Low intensity precursors will lead to noisier fragmentation spectra, which increase the chance of misidentification. The crosslinks that map to known structures (and especially self links) skew towards higher abundance proteins, especially in interactome studies and as therefore have a much lower misidentification rate than the remainder of a dataset.

Crosslinking-MS data deposition and reporting aims have recently been published, in order to establish clear standards for the integration of crosslinking-MS into the Worldwide Protein Data Bank (wwPDB) and its Integrative/Hybrid Methods database, PDB-Dev (Leitner

et al., 2020). Similar reporting standards should be adopted for the use of crosslinking-MS data in String (Szklarczyk et al., 2018) and other protein-protein interaction databases.

What is a "crosslink"? Information, representation and ambiguity in crosslinking-MS

As we have seen, what is commonly referred to as a "crosslink" or a residue-residue pair, is an aggregation of CSMs and peptide pairs derived from a sample crosslinked with a reagent of known specificity and spacer length (Fig. 2A). However, the bottom-up proteomic workflow used to identify crosslinks from biological samples introduces some ambiguity in crosslinking-MS data (Fig. 2) (Piotrowski and Sinz, 2018; Rappsilber, 2011; Shi et al., 2014).

The FDR of a given crosslink, covered above, reports on its identification ambiguity. An additional layer of ambiguity is present in systems with more than one copy of a given protein. Here, proteolytic digestion makes it impossible to know whether a peptide pair comes from a single or multiple polypeptide chains, unless a crosslink is observed between two identical or overlapping peptide pairs (Maiolica et al., 2007) or isotope labelling is employed (Lima et al., 2018). This property is sometimes referred to as "compositional ambiguity" (Fig. 2B) and reduces crosslinking-MS information in homomultimeric assemblies (Gaber et al., 2019) by making the interpretation of self crosslinks challenging.

Finally, the crosslinker kinetics and half-life introduce ambiguity in terms of the identification of the conformational state of dynamic systems (Fig. 2C). Given that the crosslinking reaction is performed in solution, highly solvent-accessible loops and disordered regions tend to be over-represented in crosslinking data, and the contribution of their multiple states may need to be disentangled. It has been noted that NHS-esters in particular, due to their specificity and long half-life, can trap minor conformations by sitting on a dynamic interface long enough for two lysine residues to come in contact with each other (Ziemianowicz et al., 2019). Dynamics of protein structure reflected in crosslinks can be addressed in multiple ways during modeling. One approach relies on explicitly modeling multi-state ensembles to account for multiple observed conformers in the crosslink data (Geng et al., 2017; Molnar et al., 2014). Alternatively, the data may be parsed modeling a single conformer but randomly discarding subsets of crosslinking-MS restraints, such as ambiguous interaction restraints in HADDOCK (Dominguez et al., 2003), or by resampling the restraint uncertainty during structure calculation (Erzberger et al., 2014). If the crosslinked are simply used to score models obtained from other techniques, normal mode analysis or molecular dynamics simulations may be used to obtain conformers that may account for the residue pairs observed in solution (Degiacomi et al., 2017).

The problem of state ambiguity may be compounded at high crosslinker concentrations by zippering effects, in which multiple crosslinker moieties on the same molecules are allowed to react and bring two proteins closer together than they would normally be (Belsom and Rappsilber, 2021). These artefacts can mislead the interpretation of dynamic equilibria in structural studies (Fig. 2C), resulting in overly compact structural models.

Despite having a more complex chemical space and a lower abundance of each crosslinked peptide pair due to their lower specificity (Conway et al., 2021; West et al., 2021), diazirine crosslinkers such as sulfo-SDA can provide a "truer" snapshot of conformations in solution because of the extremely short half-life of the carbene intermediate generated upon photo-activation. This lack of conformational trapping of diazirine crosslinkers, combined with their higher data density (Fig. 3), is contributing to making them a preferred choice for purified protein complexes with prevalent conformational dynamics (Ziemianowicz et al., 2019). Similarly, slow-reacting crosslinkers also prevent trapping of short-lived states, since reaction requires prolonged presence of the reactive residues in close proximity (Barrow et al., 2019).

However, low-specificity crosslinkers heighten another type of ambiguity: because of the often incomplete fragmentation information in spectra of crosslinked peptide pairs, it is often not possible to unambiguously determine which residue has reacted with the crosslinker. Addressing crosslink site ambiguity would require the development of appropriate scores and statistics for false localization rates in crosslinking peptide pairs, which is currently an area of active research.

Thus, different crosslinkers provide different in-solution snapshots of a sample, and the appropriate care should be taken when integrating crosslinking-MS data from different crosslinkers, or in studies incorporating data derived from multiple crosslinkers.

Crosslinking-MS in structural biology (1): visualization, model building and model validation

The capacity of crosslinking MS to provide distance restraints in complex samples, or even in cells, means that it has proven to be an ideal complementary technique for validating electron microscopy or crystallographic models in their native context (Casañal et al., 2019; O'Reilly and Rappsilber, 2018; Schmidt and Urlaub, 2017). Another main application of crosslinking-MS in the past decade has been to provide aid in manual model building or model validation for near-atomic resolution cryo-EM densities, or to verify unexpected conformations in solution. Some recent examples include resolving the subunit arrangement in the catalytically active U2/U6 RNA spliceosomal structure (Townsend et al., 2020), the different functional states of the condensin ATPase cycle (Lee et al., 2020) (Fig. 3), the validation of the structure of the centromeric nucleosome-CCAN kinetochore complex (Yan et al., 2019) and aiding model building of the human B^{act} complex (Haselbach et al., 2018).

The simplest way to visualize crosslinking-MS data in a structural context has been to map residue-residue pairs onto structures as Euclidean distances in a visualization software such as ChimeraX (Pettersen et al., 2021), PyMol (Schiffrin et al., 2020) or Chimera (Pettersen et al., 2004)— where the plugin XlinkAnalyzer provides a framework for the inspection of crosslinking-MS data (Kosinski et al., 2015). This simple approach is certainly powerful, and allows the user to observe the most salient features of the data, such as whether a protein fold or a particular protein-protein interface is broadly compatible with the detected crosslinks (Fig. 3). Additionally, specialized crosslinking-MS visualization software such as xiVIEW (Graham et al., 2019), Cross-ID (de Graaf et al., 2019) and XLinkDB (Schweppe et al., 2016) can combine structure mapping capabilities with browsing of spectra, sequence-based protein annotation and network visualizations.

Often, crosslinks are mapped as Cα-Cα distances, in order to allow for different side chain conformations in solution than in the vitrified or crystallised sample (Fig. 4A). In these cases, crosslink satisfaction is assessed on the basis of an upper distance limit, which is often set to the crosslinker spacer length plus the maximum length of the appropriate side chain rotamers and a small additional tolerance to account for molecular motions. As an example, the maximum distance for DSS and BS^3, which have 11.4 Å spacer, has often been set as high as 30 Å Cα-Cα (Lee et al., 2020; Tüting et al., 2020; Yan et al., 2019). Testing this upper bound assumption for DSS by examining lysine-lysine distances in molecular dynamics simulations and comparing them to experimentally observed crosslinks led to a good agreement based on average protein motions, supporting the incorporation of tolerances (Merkley et al., 2014).

As a useful check, the program DisVis (van Zundert and Bonvin, 2015; van Zundert et al., 2017) has been used to map the interaction volume in which the most crosslinks are satisfied, as was done for mapping interfaces in mitochondrial respiratory chain

supercomplexes (Liu et al., 2018). DisVis can also perform a bootstrap-based statistical analysis to detect crosslinks arising from a second conformation and false positives. This is particularly useful for mapping interaction volumes of flexible or disordered proteins (Niemeyer et al., 2020).

More sophisticated methods to assess models based on Euclidean distances rely on the expected distance distribution generated by a given crosslinker. However, these often require a calibration of such a distance distribution under the given experimental conditions, and may be prone to artefacts due to the varying level of dynamics between loops and secondary structure elements. Nevertheless, a crosslinker distance distribution provides a prior to which models may be compared– it is unlikely that in a dataset, all residue pairs fall right under the upper distance threshold. Expected distance distributions have also been generated from molecular dynamics simulations (Green et al., 2001), which however require parametrization of the crosslinker and/or the crosslinker-modified residues. To the best of our knowledge, there is no data modeling of distance distributions using explicit crosslinks in proteins.

While Euclidean approaches are computationally inexpensive and appropriate for simple model validation, more sophisticated crosslink representations have been developed to enable a more rigorous assessment of models based on crosslinking data. Crosslinkers require the presence of a solvent-accessible path between the two moieties being crosslinked. Thus, the solvent-accessible surface distance (SASD) between two residues should be considered when assessing crosslink violations (Fig. 4A). Calculations of such paths may be performed by the programs JWalk (Bullock et al., 2016) and XWalk (Kahraman et al., 2011). The drawback of these methods for structure assessment is their computational cost and the requirement of an atomic representation in order to properly calculate solvent accessibilities. Nevertheless, approximations may be made by using backbone-only models. A further complication is the relatively high false negative rate, if protein motions are not taken into account (Degiacomi et al., 2017).

The computationally expensive nature of SASD calculations also restricts these methods to assessing models obtained from other techniques, rather than employing SASD restraints during model calculation in integrative approaches. However, SASD methods have been incorporated into scoring functions capable of improving model selection, such as MNXL (Bullock et al., 2018) and its energy-based counterpart, NRGXL (Filella-Merce et al., 2020), which relies on simulations of the crosslinker path to derive an energy associated with each crosslink. Recently, the Topf group has developed XLM-Tools, a program to score models on a combination of crosslinker SASD and experimental detection of crosslinker-modified linear peptides (“monolinks”) into a single crosslink/monolink combined score (Sinnott et al., 2020). Interestingly, it has been shown that incorporating normal mode analysis or other approaches to generate protein motion *in silico* increases the precision and accuracy of scoring crosslinks with both SASD and NRGXL, further confirming the need to account for the in-solution nature of the crosslinking-MS measurement (Degiacomi et al., 2017; Filella-Merce et al., 2020).

Besides their use in selecting between alternative models generated by other structural approaches, such scoring methods may also be used to provide experimental validation or to select between different structural models (Bullock et al., 2016; Tüting et al., 2020). In multi-state systems, quantitative crosslinking-MS may be used to disentangle the contribution of individual states, especially in response to perturbation, as reviewed in (Chen and Rappsilber, 2018).

The high-throughput nature of crosslinking-MS, especially when coupled with protein and peptide-level fractionation, holds the promise of system-wide residue-residue interaction

maps. These may be used to select or validate experimentally-derived models. Recently, the breakthroughs in protein structure prediction by the deep learning algorithms AlphaFold2 (Jumper et al., 2021; Tunyasuvunakool et al., 2021) and RosettaFold (Baek et al., 2021). Here, crosslinking-MS may be used to understand the arrangement of multidomain proteins, and validate or select between proteins with different arrangements of low-confidence regions, as well as to provide evidence for protein complex interfaces. These breakthroughs usher in a new era in structural biology. In the short term, the new models will provide an improved starting point for integrative structural biology approaches (Masrati et al., 2021), where crosslinking-MS can provide valuable contact information.

Crosslinking-MS in structural biology (2): crosslinking restraints in sampling and integrative modeling

Large macromolecular assemblies that cannot be purified or tackled by traditional methods present one of the main challenges in structural biology. To deal with the flexibility and complexity of these systems, integrative structural biology seeks to combine information from multiple experiments to generate a model reflecting the best possible agreement of all experimental data and prior physical, experimental and statistical knowledge (Koukos and Bonvin, 2020; Rout and Sali, 2019; Shin et al., 2020) (Fig. 4).

Broadly speaking, integrative modeling is an iterative process of data acquisition, representation, sampling, scoring and model assessment (Fig. 4). Crosslinking-MS data may be used as restraints to drive sampling, as a scoring method for model selection, or both. Typically, crosslinking-MS supplements EM information in integrative modeling whenever near-atomic resolution data is not achievable (Leitner et al., 2016). This may be because of sample-specific limitations, inherent flexibility and disorder in the system. Additionally, crosslinking-MS may be used to place specific components that are missing or not resolved in cryo-EM densities. Models may be placed in electron densities using crosslinking-MS data as rigid or flexible bodies, or to resolve ambiguity at near-atomic resolution.

While such approaches can be used to combine a vast array of data, we will mainly focus on one of the most common applications, namely the combination of crosslinking-MS data with shape information from electron microscopy. Nevertheless, the availability of high-density crosslinking data has also significantly contributed to the quality of modeling approaches where crosslinking-MS is the sole experimental data source (Fig. 3). Thus, a combination of three crosslinking strategies generated sufficient data density for the modeling of the yeast proteasome regulatory particle (Mintseris and Gygi, 2020). Even higher data density was achieved using photo-crosslinking and was sufficient to determine the fold of protein domains from crosslinking-MS data and fragment libraries (Belsom et al., 2016).

Currently, the main general computational pipelines for deriving integrative models are the Integrative Modeling Platform (IMP) (Russel et al., 2012), HADDOCK (Dominguez et al., 2003) and Rosetta (Gray et al., 2003; Leaver-Fay et al., 2011; Leman et al., 2020). These pipelines are able to incorporate data from multiple experimental sources together with physical and statistical potentials to derive structural models representing the consensus of all available information with quantifiable precision. Besides these general programs, several other *ad hoc* pipelines have been constructed for handling crosslinking-MS restraints in sampling and scoring, either by adapting the existing frameworks or with fresh implementations (Bonomi and Camilloni, 2017; Ferber et al., 2016; Kalinin et al., 2012; Woetzel et al., 2011).

As one of the main problems in integrative modelling is sampling the vast conformational landscape, different compromises have been struck between ensuring a

thorough sampling and having the representation that most accurately reflects the precision of the input data and the available input models. Such compromises may include coarse graining of particles or simplified representations of computationally expensive restraints. Therefore, crosslinking-MS data from soluble crosslinkers is often assessed on the basis of Euclidean distance, rather than SASD. Similarly, gaussian mixture models have been successfully employed to represent medium- and low-resolution EM densities (Kawabata, 2008). In the modeling procedure, the ruggedness of the conformational space to be sampled is a function of both the resolution of the representation of the data and of the level of coarse graining employed to represent the components.

Integrative Modeling Platform (IMP)

IMP-based approaches can bring shape information, physical restraints, crosslinking-MS data, and data from other approaches into a single scoring function used in both sampling and model scoring/selection. Whenever possible, restraints are represented in a Bayesian framework in which each data source has fixed or sampled prior and computed posterior probabilities (Rieping et al., 2005). The forward model used in the IMP Bayesian framework for crosslinking-MS restraints includes two parameters to account for the uncertainty in the position of the crosslinked residues and the chance of a false positive crosslinks (Erzberger et al., 2014) (Fig. 4A). These parameters may be globally sampled, fixed across a crosslinking-MS dataset, or even optimised for each individual crosslink. The forward model determines the prior likelihood of observing a crosslink between two residues, and can then be used to compute the likelihood of a model given the experimental crosslinks. The model also handles compositional ambiguity (Fig. 3B), and can be used in multistate calculations to handle state ambiguity (Fig. 3C).

In a common application of IMP supported by the PMI platform (Shi et al., 2014), these approaches are combined with a coarse graining of the system and/or of the data, or even multiscale representations, in order to ensure the feasibility of complete sampling and to accurately represent uncertainty in input models. IMP then performs stochastic sampling of model configurations using for example a replica exchange Monte Carlo procedure. Among the advantages of this stochastic sampling approach lies the ability to test for sampling exhaustiveness and convergence (Viswanath et al., 2017a), and therefore derive a reliable precision estimate for the sampling procedure and an uncertainty in the selected model ensemble. This family of approaches has been used to obtain integrative models of a wide range of biological systems based on the integration of crosslinking-MS and other data sources. Thus, the model of the Nup84 complex (Shi et al., 2014) was obtained by combining crosslinking-MS information with negative stain EM projections. The human COP9 signalosome model was derived using data from multiple crosslinkers (Gutierrez et al., 2020) and the *M. pneumoniae* expressome architecture was obtained from in-cell crosslinking-MS and in-cell electron tomography data (O'Reilly et al., 2020).

The flexible nature of IMP has allowed it to integrate a very large amount of data sources in various combinations, including hydrogen-deuterium exchange mass spectrometry (Saltzberg et al., 2017), native mass spectrometry (Politis et al., 2015), yeast two-hybrid data (Viswanath et al., 2017b), and small-angle scattering (Kim et al., 2014). The ability of IMP to coarse grain the system and to hold multiple representations at the same time has made it particularly powerful as a tool to derive hybrid models when full atomic models of the components are not available. For example, crosslinking-MS restraints were used to derive the binding mode of a coarse-grained model of GDown1 against an atomic model of RNA polymerase II (Jishage et al., 2018)

Another way to ensure thorough sampling has been to employ crosslinking-MS and other restraints separately in successive modeling rounds. In such an implementation, different restraints drive sampling at each stage, or are employed separately in sampling, refining, or scoring. The IMP-based Assembline pipeline was employed to model the nuclear pore (von Appen et al., 2015; Kosinski et al., 2016) and yeast elongator complexes (Dauden et al., 2017). This pipeline begins with systematic rigid-body fitting of each component of the system in EM densities, followed by scoring of resulting models according to steric clashes, EM fit cross-correlation and crosslinking-MS distance restraints.

HADDOCK

In the HADDOCK approach, models are given a full atomic representation or a low level of coarse graining using MARTINI (Roel-Touris et al., 2019). HADDOCK has often been chosen when a significant portion of the system is covered by all-atom structures or high-quality homology models. Sampling is achieved by an initial rigid-body conformational search, followed by a round of flexible refinement and a round of refinement in explicit solvent. Minimization in each sampling round is performed with an energy function that combines the physical force field and experimental restraints (Dominguez et al., 2003; Karaca and Bonvin, 2013). At each stage, models are selected according to a HADDOCK score, which is a combination of an all-atom physical potential, buried surface area and restraint energy derived from experimental restraints. Crosslinking-MS data may be encoded as distance restraints using an upper bound harmonic function with a flat bottom potential, often between Cα atoms. However, modeling crosslinks as restraints between Cβ atoms, or between the appropriate side chain moieties greatly increases the precision of the derived models by tightening loose restraints and somewhat preventing restraint satisfaction by unrealistic crosslinks through the backbone (Gong et al., 2020).

Much like IMP, HADDOCK has also been developed to accept a wide range of experimental information which may be entered as “ambiguous interaction restraints”, including NMR chemical shift perturbation, limited proteolysis mass spectrometry (Hennig et al., 2012) and mutagenesis. Crosslinking-MS data may also be entered as unambiguous distance restraints to supplement ambiguous interaction restraints from DisVis interface mapping or from other techniques. This approach was for example employed in the modeling of the free C6 protein from the complement cascade from crosslinking-MS data (Hevler et al., 2021). These restraints are then used in sampling and scoring phases. The recent release of HADDOCK2.4 has enabled users to incorporate electron microscopy densities as an active energy term in each sampling stage (van Zundert et al., 2015), and in model scoring. Restraints may also be employed sequentially: the HADDOCK-M3 pipeline (Karaca et al., 2017) employs distance-restraint driven rigid body sampling, followed by a flexible refinement and rescoring of models according to experimental information not used in sampling, for example from solution scattering or cryo-EM. Similarly, a protocol to probe the structure of complexes at a high throughput involves iterative rounds of structure prediction, crosslinking-MS scoring, and HADDOCK-based protein-protein docking (Orbán-Németh et al., 2018). Rescoring may also be performed using SASD-based metrics (Bullock et al., 2018).

Rosetta

The powerful modeling suite Rosetta has also been employed in integrative modeling (Leman et al., 2020). Rosetta approaches can use both deterministic and stochastic sampling and rely on the satisfaction of the Rosetta energy function. This function is a linear combination of terms balancing physics-based and statistically-derived potentials (Alford et al., 2017;

Leman et al., 2020). In this framework, crosslinks have also been represented as harmonic upper bound distance restraints and employed in guiding structure prediction, comparative modeling and model validation (Kahraman et al., 2013). Another implementation used a modified Lorenz function to implement the possibility of false positive restraints at high FDR (Belsom et al., 2016). These restraints may be incorporated during sampling, or to score models derived by *ab initio* protocols such as RosettaDock (Marze et al., 2018) or trRosetta (Yang et al., 2020). The modular nature of the software has also allowed the integration of crosslinking-MS data into more complex pipelines, such as the targeted MS/crosslinking workflow used to derive a model of the interactions between human and *S. pyogenes* M1 proteins (Hauri et al., 2019). Finally, crosslinking-MS has also been used as distance restraints to supplement molecular dynamics-based folding approaches (Brodie et al., 2017).

Machine learning approaches for predicting and placing protein structures according to multiple sources of information have also shown great promise even beyond the individual protein cases tackled in CASP (Jumper et al., 2021; Kryshtafovych et al., 2019). Recently, the structure of the ~800KDa Fanconi Anemia complex was derived using Rosetta at atomic detail using a 4.6 Å resolution cryo-EM density and distance restraints derived from crosslinking-MS data and from residue-residue contacts obtained with deep convolutional neural networks (Farrell et al., 2020). Publicly available webservers such as ColabFold are beginning to use the Alphafold framwork for protein complex prediction (Mirdita et al.), expanding to incorporate experimental data (Humphreys et al., 2021).

Outcomes of integrative modeling with crosslinking-MS

A successful outcome of integrative modeling is the convergence of good-scoring models into clusters of similar solutions. Several criteria for clustering integrative models have been proposed (Karaca et al., 2017; Rodrigues et al., 2012; Viswanath et al., 2017a). Lack of convergence may be indicative of insufficient data, insufficient sampling, or not having accounted for multi-state or heterogeneous systems. For this reason, data collection, model representation and modeling is an iterative process (Fig.4).

Ensembles of solutions derived from integrative modeling procedures must have a quantifiable model precision (Schneidman-Duhovny et al., 2014). This is often stated as the rmsd from the centroid structure of the cluster of solutions being selected. This model precision measure is effectively a measure of the uncertainty derived from experimental data and the modeling procedure itself. In the case of crosslinking-MS data, the quantified experimental uncertainty rests on accurate estimation of FDR at the correct level of analysis. The model precision estimate also relies on having sufficiently sampled the conformational space at a given sampling precision.

Integrative models or ensembles coming out of the modeling procedure are deposited in the PDB-Dev repository in IHM format (Burley et al., 2017; Vallat et al., 2019). This dictionary-based format can be visualized in ChimeraX (Pettersen et al., 2021), and can provide full reporting of building blocks, modeling procedure, restraints, model precision and localization probability densities by directly pulling information and data from relevant databases (Berman et al., 2019).

Crosslinking-MS in biological discovery: *in situ* structural interactomics

Crosslinking-MS has matured into a technology to map protein-protein interactions at a high throughput, in near-native conditions, and without protein tagging. It reports binary

interactions, which is an advantage over traditional affinity purification-based or co-fractionation protein interactome mapping studies. The additional residue-residue pair information of crosslinking-MS PPI networks allows mapping of interfaces between partner proteins (Fig. 5).

In large-scale crosslinking-MS studies, the crosslinking reaction is performed directly on a complex sample, resulting in a “snapshot” of all protein-protein interactions (and of residue-residue distances within proteins) captured by the crosslinker. Moreover, in-cell or in-organelle crosslinking-MS studies have to be performed with crosslinkers capable of crossing cell membranes, such as DSSO and DSS. Interestingly, the recent development of antibodies to detect DSSO and DSBU-crosslinked peptides in immunostaining has shown that at low concentrations the crosslinker forms a gradient across HEK293 cells (Singh et al., 2021). Thus, just like in the studies of purified systems, in-cell crosslinking-MS studies require a careful optimisation of reaction conditions to balance between sufficiently high levels of crosslinking and the potential introduction of artefacts by zippering effects or the formation of higher-order crosslinked aggregates. As for structural modeling, the interpretation of crosslinking-MS PPI networks rests on the establishment of a reliable FDR at the level of interpretation, i.e. the PPI level. Currently, comparison between datasets is somewhat hampered by a current lack of an agreed error estimation procedure at the PPI level in the crosslinking-MS field. Together with deposition of the data and other measures of standardisation this has been recognised by the field and first steps to address this have been taken (Leitner et al., 2020).

The challenges presented by sample complexity and the huge span of protein abundances have then to be met by multiple rounds of chromatographic enrichment of crosslinked peptides and/or by additional biochemical purifications. The dynamic range of native protein abundances is especially problematic: as crosslinking reactions can be approximated by second-order kinetics, there is a strong bias towards crosslinking the most abundant proteins (Fürsch et al., 2020).

Thus, protein and peptide-level fractionation are particularly critical to achieve the restraint density required by the use of crosslinks as restraints in integrative structure modeling, especially on low-abundant proteins. While multidimensional fractionation has pushed the limit of tag-free chromatographic enrichment, we envision that new generation enrichable crosslinkers based on previously underutilized chemistries (Rey et al., 2021; Stadlmeier et al., 2020) will provide opportunities for the selective enrichment of crosslinked peptides.

The PPI network resulting from large-scale studies reports on pairwise interactions of proteins, meaning each edge represents a contact at the resolution of the crosslinker spacer, given protein motions in solution. This is different from PPI networks inferred from coexpression, thermal profile association (Mateus et al., 2020), or pulldown approaches, which report on correlated behavior under the given assay condition and not on direct through-space interactions (Low et al., 2021). However, the pairwise nature of the network is an inherent limitation in proving the composition of multimeric assemblies especially in cases comprising uncharacterised proteins. Co-fractionation-MS experiments (Skinnider and Foster, 2021) have been successfully employed to impute higher-order interactions through co-elution of proteins over a size exclusion column, with direct interactions verified by crosslinking data (Kastritis et al., 2017).

Several labs have focussed on developing a combination of pulldowns and crosslinking-MS (Herzog et al., 2012; Liu et al., 2017b; Makowski et al., 2016). In these workflows, crosslinking may be performed in the original sample, however complex, in order

to probe all available interactions, or on-bead (Liu et al., 2017b; Makowski et al., 2016). On-bead crosslinking increases the signal-to-noise of the experiment by providing a less complex sample to the crosslinking reaction and to the MS, but by necessity will yield only high-affinity interactions that are preserved through washing steps. On the other hand, in-cell crosslinking followed by affinity purification can yield native protein-protein interactions (Slavin et al., 2021), but needs to be carefully optimised, especially if the protein of interest is a binder of highly abundant proteins or proteins that may interact with the cytoskeleton.

Considerable depth may also be achieved by crosslinking subcellular fractions. In a systematic crosslinking-MS study in mouse synaptic vesicles, the authors employed sequential digestion and SCX fractionation to enrich for crosslinked peptides, yielding 11,999 unique residue-residue pairs involving 2362 proteins (Gonzalez-Lozano et al., 2020). The restraints obtained by crosslinking-MS confirmed known interactions and provided new insights into the arrangement of synaptic SNARE proteins. Moreover, crosslinking-MS restraints mapped the binding sites of several novel interactors onto the crucial AMPAR ligand-gated ion channel, responsible for mediating fast neurotransmission. Similarly, mitochondrial crosslinking-MS PPI networks have detected interactions between complexes in the respiratory chain, proposing arrangements for an elusive respiratory chain supercomplex (Liu et al., 2018; Ryl et al., 2020; Schweppe et al., 2017).

In bacteria, current approaches have proven sufficient to provide a good proteome coverage in crosslinking-MS data. Thus, whole cell studies from in-cell crosslinking of *M. pneumoniae* performed with DSS, DSSO, multidimensional fractionation and sequential digestion detected 1957 heteromeric crosslinks (O'Reilly et al., 2020) (Fig. 5A). The PPI network allowed the authors to confirm subunit arrangements inferred by homology in known complexes (Fig. 5B) and to annotate a novel RNAse Y complex. Multiple residue-residue pairs mapped interactions between the universal transcription termination/antitermination factor NusA, the RNA polymerase and the ribosome. These findings provided a starting point for a systematic in-cell cryo-ET investigation, which verified the ternary nature of the complex inside cells. The in-cell crosslinking-MS and cryo-ET data were then combined to derive an integrative model for the expressome complex comprising NusA, RNA polymerase, NusG and ribosome, revealing a novel arrangement of the components present inside cells, as well as unique aspects of the *M. pneumoniae* ribosome.

Outlook: in-cell structural biology

The growing application of crosslinking-MS as a system-wide interactome mapping technology also provides an avenue for interactome screening studies, in which pairwise protein interactions are probed at a system-wide scale without the need for protein tagging. The PPIs revealed in crosslinking-MS studies come with a quantified FDR and can provide a basis for further biochemical, structural and functional experiments, including quantitative proteomic approaches. In this regard, crosslinking-MS can fill a niche and supplement interactome mapping approaches for RNA-protein, DNA-protein and metabolite-protein interactions in understanding the key players and mechanisms involved in cellular processes.

The unique combination of PPI mapping and high-throughput, spatial restraints provided by crosslinking-MS have been particularly beneficial to the interpretation of low, medium-resolution and near atomic-resolution EM densities, which have revolutionised structural biology over the past decade. The further development of in-cell crosslinking-MS workflows and the increased resolution of cryo-ET is also ushering in an era of structural studies directly performed *in situ*. Alongside this, additional techniques, such as in-cell NMR (Freedberg and Selenko, 2014; Luchinat and Banci, 2017) and other structural proteomics

approaches such as limited proteolysis mass spectrometry (Schopper et al., 2017) are also providing critical insight into biological structures *in situ.*

The constant development of new approaches in both the EM and crosslinking-MS fields are unlocking the ability to study deeper fractions of the proteome at previously unattainable resolutions. The improvement of 3D classification has allowed EM to characterise transient or weak interactions, time-resolved phenomena, or the quantitative effects of perturbations. For crosslinking-MS, understanding these phenomena has required improvements in data amount, data density and crosslink quantitation. The employment of these technologies in characterising dynamics *in situ* will necessarily need to be complemented by molecular dynamics and other simulation techniques in integrative approaches.

At the same time, the revolutionary developments in the field of structure prediction (Baek et al., 2021; Jumper et al., 2021; Tunyasuvunakool et al., 2021) have also opened up the possibility of system-wide structural interactome studies using newly derived predictions for previously uncharacterised proteins. At the very least, these models have provided an improved starting point over previous *ab initio* and comparative modeling approaches that were employed in cases of poor structural coverage, though the uncertainty associated with models predicted by deep learning approaches remains to be rigorously established.

These developments push for the need of integration of structural data into whole-cell modeling approaches, providing spatial and molecular resolution to systems biology. This information can lead to a structural characterisation and modeling of entire biological processes, which will require integration of structural, kinetic and -omics data in whole-cell quantitative models (Earnest et al., 2018; Feig and Sugita, 2019; Karr et al., 2012; Raveh et al., 2021; Singla et al., 2018).

Acknowledgments

We are grateful to Dr. Francis O' Reilly, Dr. Ezgi Karaca and Swantje Lenz for critical reading of the manuscript. **Funding:** This work received funding from the Deutsche Forschungsgemeinschaft under Germany's Excellence Strategy – EXC 2008/1 – 390540038, project no. 426290502 and 392923329, the Wellcome Trust Senior Research Fellowship (103139) to J.R..The Wellcome Centre for Cell Biology is supported by core funding from the Wellcome Trust (203149).

Figures and tables

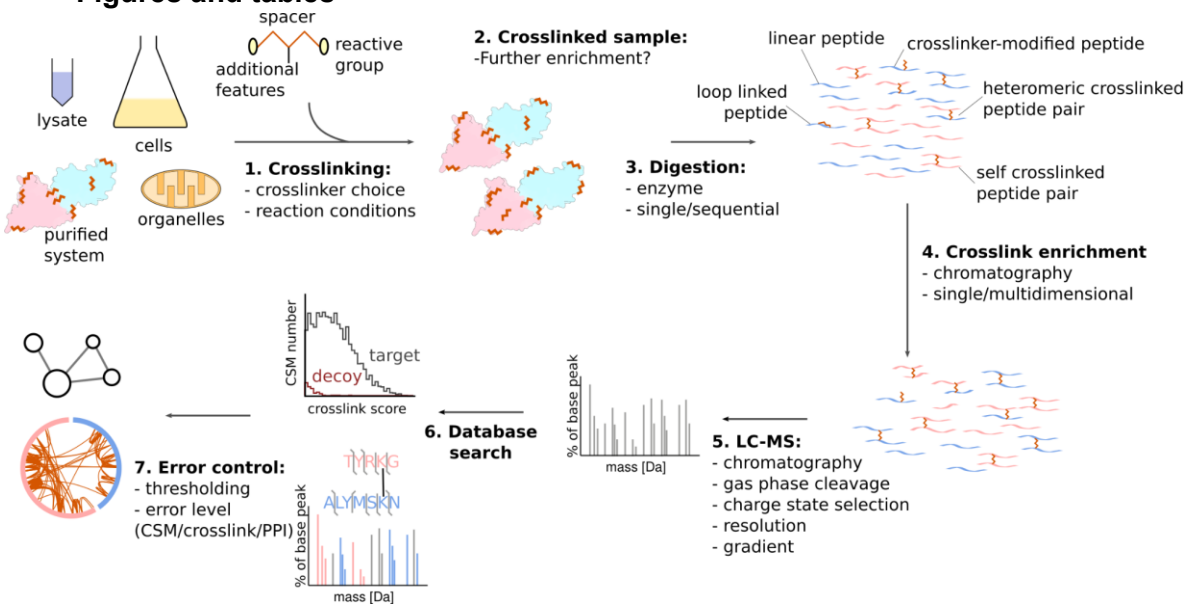

Figure 1. crosslinking-MS workflow and experimental design.

Samples of varying degrees of complexity are reacted with soluble crosslinkers. These introduce covalent bonds between residues that are close in space. The choice of crosslinker (1) should be based upon the biological question to be addressed and the complexity of the sample. After crosslinking, the sample may be further purified biochemically (2) or directly subjected to digestion (3). Sequential or parallel digestion strategies may be employed for more complex samples. The resulting peptide mixture contains crosslinked peptide pairs, loop-linked peptides (where the crosslinker reacted within the peptide), crosslinker-modified peptides and linear peptides, in order of abundance. Crosslinked peptides may then be enriched by a single or multiple rounds of chromatography, which may comprise SEC, SCX, hSAX and affinity chromatography (4). The resulting samples are acquired by LC-MS using specific methods designed to detect and fragment crosslinked peptide pairs (5) and computationally searched and matched against a sequence database, yielding crosslink spectra matches (CSMs) (6). Finally, the appropriate error control is applied to derive a false discovery rate for the dataset at the level of CSMs, residue pairs, or protein-protein interactions (7).

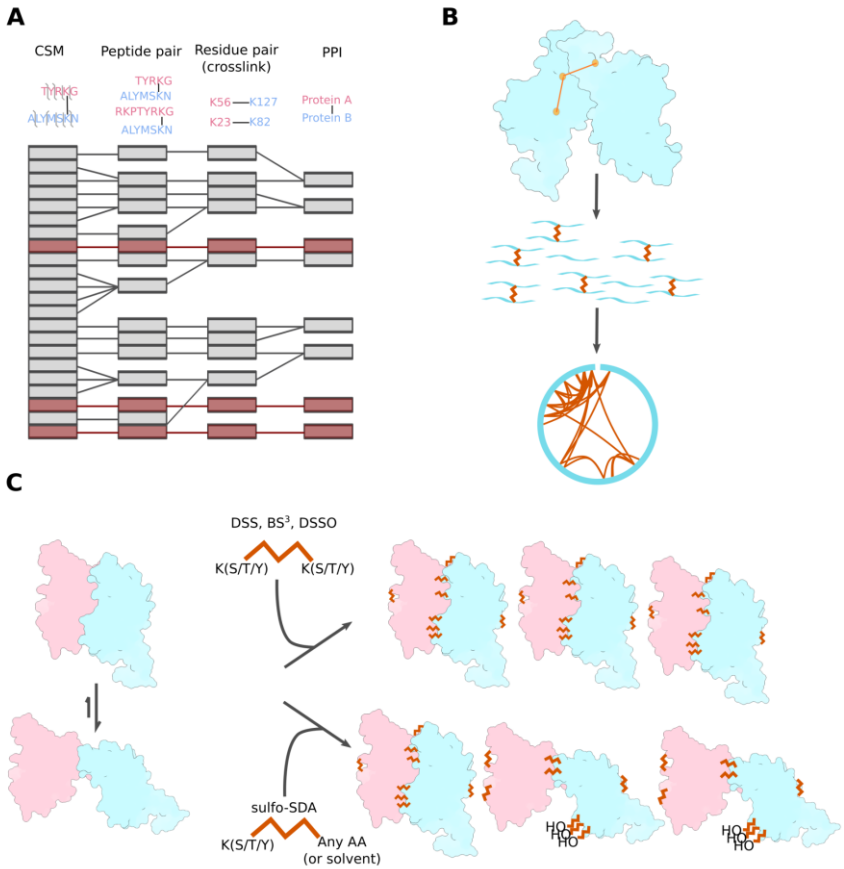

Figure 2. Error propagation and ambiguity in crosslinking-MS.
A. Identification ambiguity and error propagation. False discovery rate (FDR) thresholding at the CSM level leads to error propagation to higher levels. A set of crosslink spectra matches (CSMs) at a given FDR threshold contains target matches (grey) and decoy matches (red). Correct CSMs tend to aggregate and support correct peptide and residue pairs. As decoy matches are random, these are less likely to aggregate into peptide pairs, residue pairs ("crosslinks") and protein-protein interactions (PPIs). Thus, setting a simple 5% FDR threshold at the CSM level results in higher error rates at higher levels. Therefore, target-decoy FDR error control should be performed at the level of downstream analysis (often residue pair or PPI). **B.** Composition ambiguity. Crosslinking a homomultimeric assembly leads to loss of information upon proteolytic digestion. It is no longer possible to distinguish whether a peptide pair came from a crosslink within or between subunits sharing the same sequence, unless the same or overlapping peptide makes up a crosslinked peptide pair. This ambiguity may be treated by isotope labelling approaches. **C.** State ambiguity. The in-solution nature of crosslinking leads to snapshots of dynamic systems. Because of their long half-life and high reactivity, NHS-esters can lead to conformational trapping and the over-representation of minor conformations. Photocrosslinking with diazirine chemistry leads to highly reactive intermediates that can crosslink to protein or solvent, leading to a truer snapshot of the equilibrium in solution.

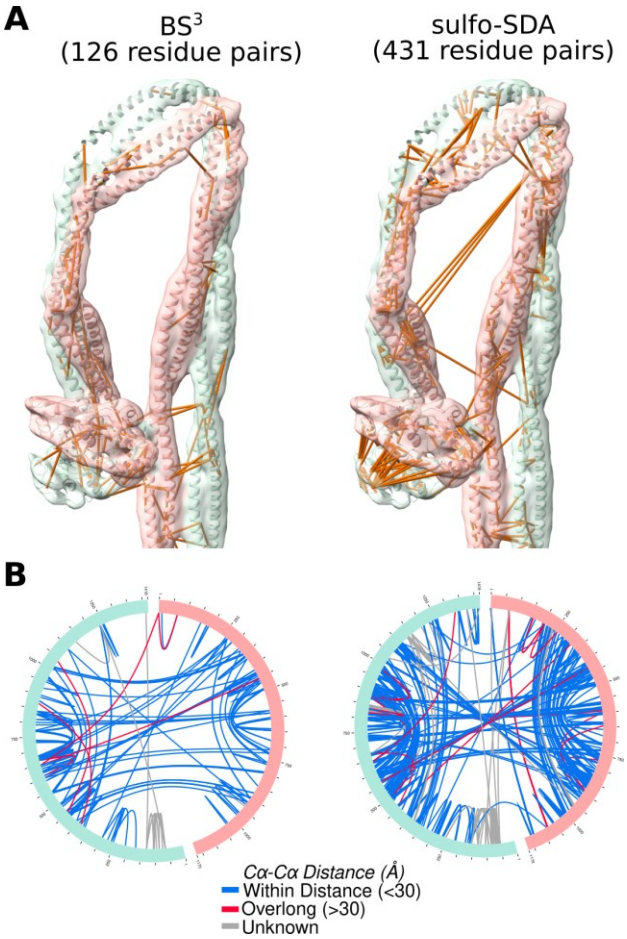

Figure 3. Crosslinking-MS for aiding model-building.
Crosslinking-MS experiments aided model-building in cryo-EM densities for the *S. cerevisiae* holo-condensin complex. Residue-residue pairs were used to guide placement of helical registers in low-resolution cryo-EM maps. Top: crosslinks (orange) detected using the NHS-ester crosslinker BS^3 and the diazirine-based crosslinker sulfo-SDA. The high density SDA data provided redundant information that guided model building of long helical regions for the smc4 (light blue) and smc2 (pink) subunits and confirmed the interface between the two subunits at the hinge. Some long-distance links capture the inherent flexibility of the large arm region. **B.** Circle view displaying self and heteromeric crosslinks of smc4 and smc2, colored according to Cα-Cα Euclidean distance..

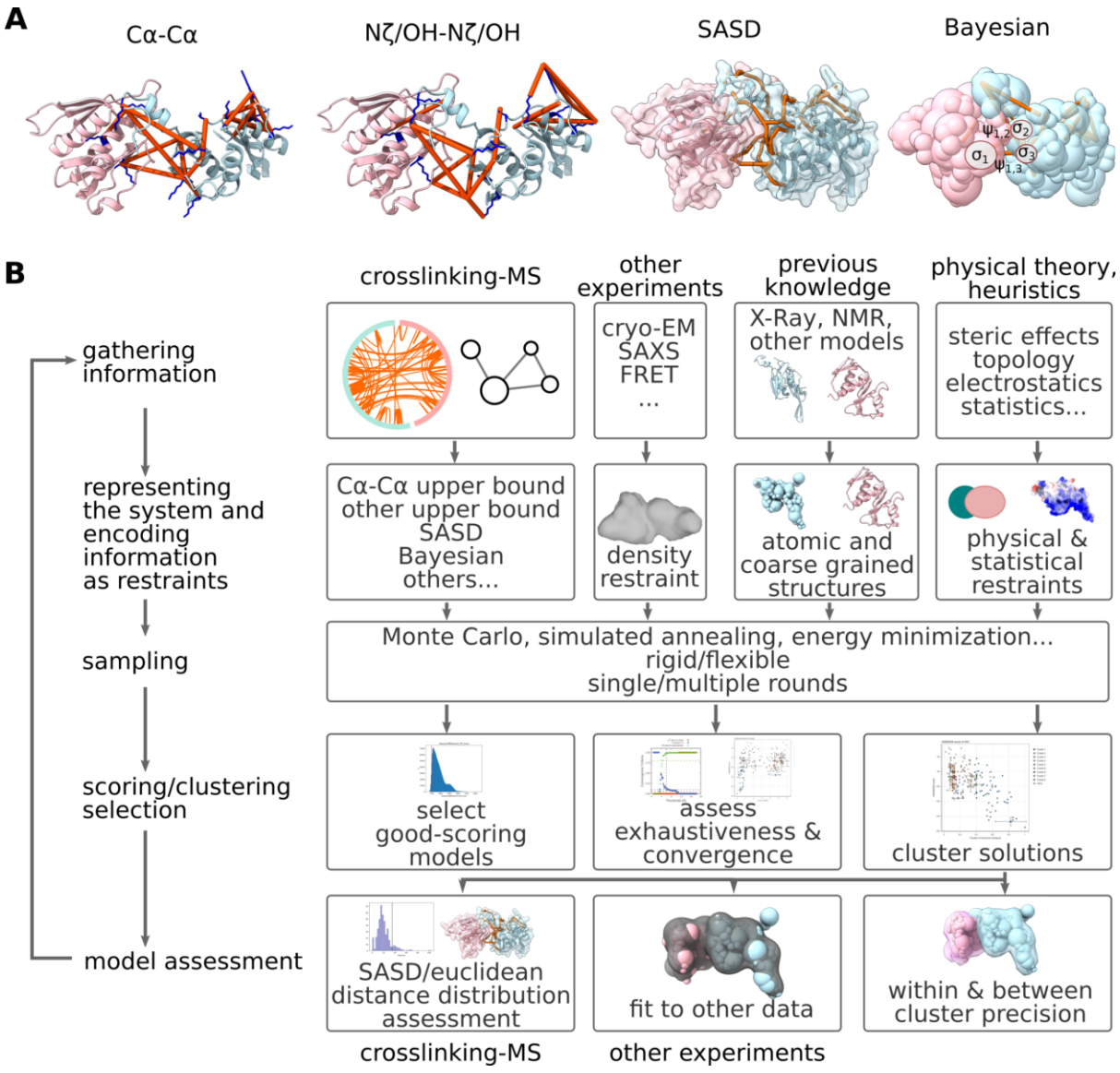

Figure 4. Crosslinking-MS data in integrative modeling.
A. Common representations of crosslinking-MS data mapped onto structures. From left to right: Cα-Cα Euclidean distance between crosslinked residues (blue), Euclidean distance between reactive groups, solvent-accessible surface distance (SASD, here calculated with Jwalk (Bullock et al., 2016)) and the Bayesian forward model to represent crosslink likelihood (Erzberger et al., 2014). In this model, the likelihood to observe a crosslink between amino acids 1 and 2 depends on the false-positive rate $\psi_{1,2}$ and the positional uncertainty of each atom σ_1 and σ_2. For crosslink visualization or model scoring, both Euclidean and SASD representations have been employed, with SASD providing improved accuracy. In integrative modeling, the first two representations are often used in HADDOCK and Rosetta, while the Bayesian representation is implemented in the Integrative Modeling Platform (IMP). **B.** A generic workflow scheme for integrative modeling and protein-protein docking using crosslinking-MS data in combination with other experimental techniques. Programs such as the IMP and HADDOCK can derive integrative models from a combination of atomic or coarse grained models, physical restraints and experimental restraints from various techniques. Typically, crosslinking-MS data is encoded by harmonic upper bound restraints or using the Bayesian approach. Appropriate quantification of uncertainty in the experimental data, starting models and modeling procedure allows for quantification of model and sampling precision. Model assessment and validation should also be performed using data not employed in the modeling procedure. For an in-depth treatment of integrative modeling approaches, see (Koukos and Bonvin, 2020; Rout and Sali, 2019; Shin et al., 2020).

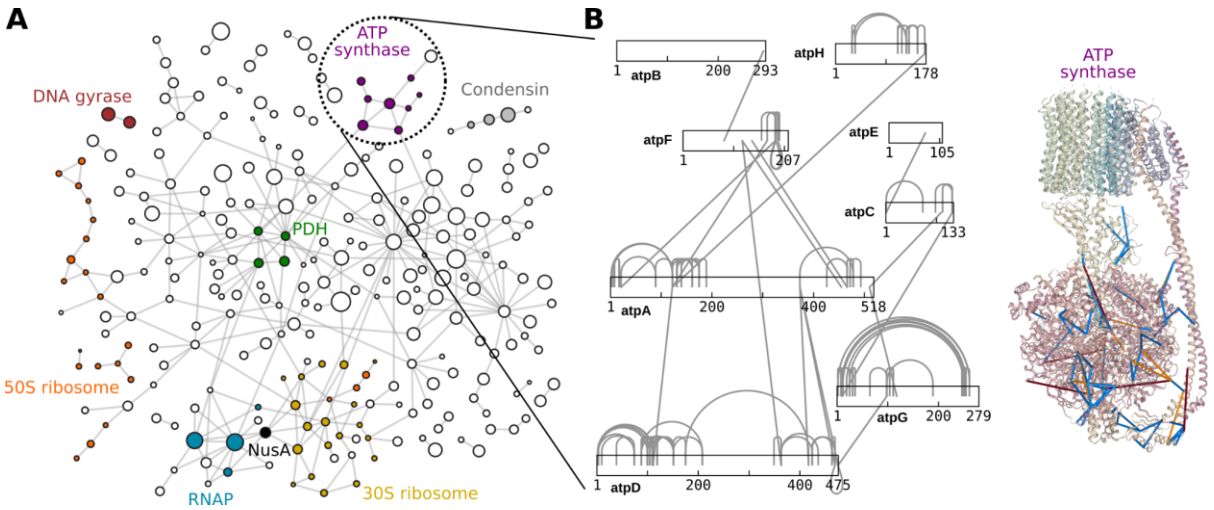

Figure 5. In-cell crosslinking-MS and PPI networks.

A. The crosslinking-MS network of *M. pneumoniae* derived from in-cell crosslinking using DSS and DSSO (O'Reilly et al., 2020). Each node represents a protein, with the size of the circle proportional to protein length. Each edge represents an interaction detected with at least a single residue-residue pair. The network has an FDR of 5% at the residue pair and PPI levels. It captures through-space interactions of several known soluble and membrane complexes and their regulatory subunits including uncharacterised binders. Moreover, it reveals key proteins, such as NusA, that link separate biological processes by direct interaction. **B.** The PPIs revealed by the network in (A) are made up of one or more residue-residue pairs, which can be used to validate models, map binding interfaces, or derive new models by docking and integrative modeling. Validation of the homology model of *M. pneumoniae* ATP synthase from the in-cell crosslinking-MS data. Crosslinks are colored by distance (blue: shorter than 28Å; yellow: between 28 and 35Å; red:longer than 35Å).

Crosslinker	Chemistry	Crosslinked residues	Spacer length (Å)	MS Cleavable?	Enrichable?	Approx. Reaction rate	Approx. Half-life	Ref.
DSS/BS^3	NHS-ester	K (STY)-K (STY)	11.4	no	no	Medium	Long (minutes-hours)	-
DSSO	NHS-ester	K (STY)-K (STY)	10.3	yes	no	Medium	Long (minutes-hours)	(Kao et al., 2011)
DSBU	NHS-ester	K (STY)-K (STY)	12.5	yes	no	Medium	Long (minutes-hours)	(Müller et al., 2010)
EDC	Carbodiimide	ED-K (STY)	"Zero length"	no	no	medium	long	-
sulfo-SDA	NHS-ester and diazirine	K (STY)-all AAs	3.4	no	no	very fast	Extremely short (nanoseconds)	(Belsom et al., 2016)
PIR	NHS-ester	K (STY)-K (STY)	~60	yes	yes	Medium	Long (minutes-hours)	(Tang and Bruce, 2010)
NHSF	NHS ester-sulfur fluoride exchange	K (STY)-STY (HK)	5.7	no	no	slow	Long (minutes-hours)	(Yang et al., 2018)

Table 1. Commonly used crosslinkers. See (Anjaneyulu and Staros, 1987) for NHS-ester kinetics, (Wang et al., 2006) for diazirine kinetics. Carbodiimide reactivity is analyzed in (Cammarata et al., 2015). Sulfonyl fluorides are described in (James, 1978) and (Lively and Powers, 1978) and reviewed in (Narayanan and Jones, 2015). As reaction rates are a function of concentration - therefore the terms here refer to the relative magnitude of the rate constants of these reagents. Similarly, half-lives are also a function of pH and temperature, especially for hydrolyzable moieties such as esters.

Agafonov, D.E., Kastner, B., Dybkov, O., Hofele, R.V., Liu, W.-T., Urlaub, H., Lührmann, R., and Stark, H. (2016). Molecular architecture of the human U4/U6.U5 tri-snRNP. Science *351*, 1416–1420.

Alber, F., Dokudovskaya, S., Veenhoff, L.M., Zhang, W., Kipper, J., Devos, D., Suprapto, A., Karni-Schmidt, O., Williams, R., Chait, B.T., et al. (2007). Determining the architectures of macromolecular assemblies. Nature *450*, 683–694.

Alford, R.F., Leaver-Fay, A., Jeliazkov, J.R., O'Meara, M.J., DiMaio, F.P., Park, H., Shapovalov, M.V., Renfrew, P.D., Mulligan, V.K., Kappel, K., et al. (2017). The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. J. Chem. Theory Comput. *13*, 3031–3048.

Anjaneyulu, P.S., and Staros, J.V. (1987). Reactions of N-hydroxysulfosuccinimide active esters. Int. J. Pept. Protein Res. *30*, 117–124.

von Appen, A., Kosinski, J., Sparks, L., Ori, A., DiGuilio, A.L., Vollmer, B., Mackmull, M.-T., Banterle, N., Parca, L., Kastritis, P., et al. (2015). In situ structural analysis of the human nuclear pore complex. Nature *526*, 140–143.

Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G.R., Wang, J., Cong, Q., Kinch, L.N., Schaeffer, R.D., et al. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. Science.

Barrow, A.S., Smedley, C.J., Zheng, Q., Li, S., Dong, J., and Moses, J.E. (2019). The growing applications of SuFEx click chemistry. Chem. Soc. Rev. *48*, 4731–4758.

Belsom, A., and Rappsilber, J. (2021). Anatomy of a crosslinker. Curr. Opin. Chem. Biol. *60*, 39–46.

Belsom, A., Schneider, M., Fischer, L., Brock, O., and Rappsilber, J. (2016). Serum Albumin Domain Structures in Human Blood Serum by Mass Spectrometry and Computational Biology. Mol. Cell. Proteomics *15*, 1105–1116.

Berman, H.M., Adams, P.D., Bonvin, A.A., Burley, S.K., Carragher, B., Chiu, W., DiMaio, F., Ferrin, T.E., Gabanyi, M.J., Goddard, T.D., et al. (2019). Federating Structural Models and Data: Outcomes from A Workshop on Archiving Integrative Structures. Structure *27*, 1745–1759.

Bonomi, M., and Camilloni, C. (2017). Integrative structural and dynamical biology with PLUMED-ISDB. Bioinformatics *33*, 3999–4000.

Brodie, N.I., Popov, K.I., Petrotchenko, E.V., Dokholyan, N.V., and Borchers, C.H. (2017). Solving protein structures using short-distance cross-linking constraints as a guide for discrete molecular dynamics simulations. Sci Adv *3*, e1700479.

Bullock, J.M.A., Schwab, J., Thalassinos, K., and Topf, M. (2016). The Importance of Non-accessible Crosslinks and Solvent Accessible Surface Distance in Modeling Proteins with Restraints From Crosslinking Mass Spectrometry. Mol. Cell. Proteomics *15*, 2491–2500.

Bullock, J.M.A., Sen, N., Thalassinos, K., and Topf, M. (2018). Modeling Protein Complexes Using Restraints from Crosslinking Mass Spectrometry. Structure *26*, 1015–1024.e2.

Burley, S.K., Kurisu, G., Markley, J.L., Nakamura, H., Velankar, S., Berman, H.M., Sali, A.,

Schwede, T., and Trewhella, J. (2017). PDB-Dev: a Prototype System for Depositing Integrative/Hybrid Structural Models. Structure *25*, 1317–1318.

Cammarata, C.R., Hughes, M.E., and Ofner, C.M., 3rd (2015). Carbodiimide induced cross-linking, ligand addition, and degradation in gelatin. Mol. Pharm. *12*, 783–793.

Casañal, A., Shakeel, S., and Passmore, L.A. (2019). Interpretation of medium resolution cryoEM maps of multi-protein complexes. Curr. Opin. Struct. Biol. *58*, 166–174.

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

Chavez, J.D., Schweppe, D.K., Eng, J.K., Zheng, C., Taipale, A., Zhang, Y., Takara, K., and Bruce, J.E. (2015). Quantitative interactome analysis reveals a chemoresistant edgotype. Nat. Commun. *6*, 7928.

Chavez, J.D., Schweppe, D.K., Eng, J.K., and Bruce, J.E. (2016). In Vivo Conformational Dynamics of Hsp90 and Its Interactors. Cell Chem Biol *23*, 716–726.

Chavez, J.D., Lee, C.F., Caudal, A., Keller, A., Tian, R., and Bruce, J.E. (2018). Chemical Crosslinking Mass Spectrometry Analysis of Protein Conformations and Supercomplexes in Heart Tissue. Cell Syst *6*, 136–141.e5.

Chavez, J.D., Mohr, J.P., Mathay, M., Zhong, X., Keller, A., and Bruce, J.E. (2019). Systems structural biology measurements by in vivo cross-linking with mass spectrometry. Nat. Protoc. *14*, 2318–2343.

Chen, Z.A., and Rappsilber, J. (2018). Protein Dynamics in Solution by Quantitative Crosslinking/Mass Spectrometry. Trends Biochem. Sci. *43*, 908–920.

Chen, Z., Fischer, L., Tahir, S., Bukowski-Wills, J.-C., Barlow, P., and Rappsilber, J. (2016). Quantitative cross-linking/mass spectrometry reveals subtle protein conformational changes. Wellcome Open Res *1*, 5.

Chen, Z.A., Jawhari, A., Fischer, L., Buchen, C., Tahir, S., Kamenski, T., Rasmussen, M., Lariviere, L., Bukowski-Wills, J.-C., Nilges, M., et al. (2010). Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry. EMBO J. *29*, 717–726.

Chen, Z.-L., Meng, J.-M., Cao, Y., Yin, J.-L., Fang, R.-Q., Fan, S.-B., Liu, C., Zeng, W.-F., Ding, Y.-H., Tan, D., et al. (2019). A high-speed search engine pLink 2 with systematic evaluation for proteome-scale identification of cross-linked peptides. Nat. Commun. *10*, 3404.

Conway, L.P., Jadhav, A.M., Homan, R.A., Li, W., Rubiano, J.S., Hawkins, R., Lawrence, R.M., and Parker, C.G. (2021). Evaluation of fully-functionalized diazirine tags for chemical proteomic applications. Chem. Sci. *12*, 7839–7847.

Dau, T., Gupta, K., Berger, I., and Rappsilber, J. (2019). Sequential Digestion with Trypsin and Elastase in Cross-Linking Mass Spectrometry. Anal. Chem. *91*, 4472–4478.

Dauden, M.I., Kosinski, J., Kolaj-Robin, O., Desfosses, A., Ori, A., Faux, C., Hoffmann, N.A., Onuma, O.F., Breunig, K.D., Beck, M., et al. (2017). Architecture of the yeast Elongator complex. EMBO Rep. *18*, 264–279.

Degiacomi, M.T., Schmidt, C., Baldwin, A.J., and Benesch, J.L.P. (2017). Accommodating Protein Dynamics in the Modeling of Chemical Crosslinks. Structure *25*, 1751–1757.e5.

Dominguez, C., Boelens, R., and Bonvin, A.M.J.J. (2003). HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. J. Am. Chem. Soc. *125*, 1731–1737.

Dorn, G., Leitner, A., Boudet, J., Campagne, S., von Schroetter, C., Moursy, A., Aebersold, R., and Allain, F.H.-T. (2017). Structural modeling of protein-RNA complexes using crosslinking of segmentally isotope-labeled RNA and MS/MS. Nat. Methods *14*, 487–490.

Earnest, T.M., Cole, J.A., and Luthey-Schulten, Z. (2018). Simulating biological processes: stochastic physics from whole cells to colonies. Rep. Prog. Phys. *81*, 052601.

Erzberger, J.P., Stengel, F., Pellarin, R., Zhang, S., Schaefer, T., Aylett, C.H.S., Cimermančič, P., Boehringer, D., Sali, A., Aebersold, R., et al. (2014). Molecular Architecture of the 40S· eIF1· eIF3 Translation Initiation Complex. Cell *159*, 1227–1228.

Farrell, D.P., Anishchenko, I., Shakeel, S., Lauko, A., Passmore, L.A., Baker, D., and DiMaio, F. (2020). Deep learning enables the atomic structure determination of the Fanconi Anemia core complex from cryoEM. IUCrJ *7*, 881–892.

Feig, M., and Sugita, Y. (2019). Whole-Cell Models and Simulations in Molecular Detail. Annu. Rev. Cell Dev. Biol. *35*, 191–211.

Ferber, M., Kosinski, J., Ori, A., Rashid, U.J., Moreno-Morcillo, M., Simon, B., Bouvier, G., Batista, P.R., Müller, C.W., Beck, M., et al. (2016). Automated structure modeling of large protein assemblies using crosslinks as distance restraints. Nat. Methods *13*, 515–520.

Fernandez-Leiro, R., and Scheres, S.H.W. (2016). Unravelling biological macromolecules with cryo-electron microscopy. Nature *537*, 339–346.

Filella-Merce, I., Bardiaux, B., Nilges, M., and Bouvier, G. (2020). Quantitative Structural Interpretation of Protein Crosslinks. Structure *28*, 75–82.e4.

Fischer, L., and Rappsilber, J. (2017). Quirks of Error Estimation in Cross-Linking/Mass Spectrometry. Anal. Chem. *89*, 3829–3833.

Fischer, L., Chen, Z.A., and Rappsilber, J. (2013). Quantitative cross-linking/mass spectrometry using isotope-labelled cross-linkers. J. Proteomics *88*, 120–128.

Freedberg, D.I., and Selenko, P. (2014). Live cell NMR. Annu. Rev. Biophys. *43*, 171–192.

Fritzsche, R., Ihling, C.H., Götze, M., and Sinz, A. (2012). Optimizing the enrichment of cross-linked products for mass spectrometric protein analysis. Rapid Commun. Mass Spectrom. *26*, 653–658.

Fürsch, J., Kammer, K.-M., Kreft, S.G., Beck, M., and Stengel, F. (2020). Proteome-Wide Structural Probing of Low-Abundant Protein Interactions by Cross-Linking Mass Spectrometry. Anal. Chem. *92*, 4016–4022.

Gaber, A., Gunčar, G., and Pavšič, M. (2019). Proper evaluation of chemical cross-linking-based spatial restraints improves the precision of modeling homo-oligomeric protein complexes. BMC Bioinformatics *20*, 464.

Geng, C., Narasimhan, S., Rodrigues, J.P.G.L.M., and Bonvin, A.M.J.J. (2017). Information-Driven, Ensemble Flexible Peptide Docking Using HADDOCK. Methods Mol. Biol. *1561*, 109–138.

Giese, S.H., Belsom, A., Sinn, L., Fischer, L., and Rappsilber, J. (2019). Noncovalently

Associated Peptides Observed during Liquid Chromatography-Mass Spectrometry and Their Effect on Cross-Link Analyses. Anal. Chem. *91*, 2678–2685.

Gong, Z., Ye, S.-X., and Tang, C. (2020). Tightening the Crosslinking Distance Restraints for Better Resolution of Protein Structure and Dynamics. Structure *28*, 1160–1167.e3.

Gonzalez-Lozano, M.A., Koopmans, F., Sullivan, P.F., Protze, J., Krause, G., Verhage, M., Li, K.W., Liu, F., and Smit, A.B. (2020). Stitching the synapse: Cross-linking mass spectrometry into resolving synaptic protein interactions. Science Advances *6*, eaax5783.

Götze, M., Iacobucci, C., Ihling, C.H., and Sinz, A. (2019). A Simple Cross-Linking/Mass Spectrometry Workflow for Studying System-wide Protein Interactions. Anal. Chem. *91*, 10236–10244.

de Graaf, S.C., Klykov, O., van den Toorn, H., and Scheltema, R.A. (2019). Cross-ID: Analysis and Visualization of Complex XL–MS-Driven Protein Interaction Networks. J. Proteome Res. *18*, 642–651.

Graham, M., Combe, C., Kolbowski, L., and Rappsilber, J. (2019). xiView: A common platform for the downstream analysis of Crosslinking Mass Spectrometry data.

Gray, J.J., Moughon, S., Wang, C., Schueler-Furman, O., Kuhlman, B., Rohl, C.A., and Baker, D. (2003). Protein–Protein Docking with Simultaneous Optimization of Rigid-body Displacement and Side-chain Conformations. Journal of Molecular Biology *331*, 281–299.

Green, N.S., Reisler, E., and Houk, K.N. (2001). Quantitative evaluation of the lengths of homobifunctional protein cross-linking reagents used as molecular rulers. Protein Sci. *10*, 1293.

Gutierrez, C., Chemmama, I.E., Mao, H., Yu, C., Echeverria, I., Block, S.A., Rychnovsky, S.D., Zheng, N., Sali, A., and Huang, L. (2020). Structural dynamics of the human COP9 signalosome revealed by cross-linking mass spectrometry and integrative modeling. Proc. Natl. Acad. Sci. U. S. A. *117*, 4088–4098.

Gutierrez, C., Salituro, L.J., Yu, C., Wang, X., DePeter, S.F., Rychnovsky, S.D., and Huang, L. (2021). Enabling Photoactivated Cross-Linking Mass Spectrometric Analysis of Protein Complexes by Novel MS-Cleavable Cross-Linkers. Mol. Cell. Proteomics *20*, 100084.

Haselbach, D., Komarov, I., Agafonov, D.E., Hartmuth, K., Graf, B., Dybkov, O., Urlaub, H., Kastner, B., Lührmann, R., and Stark, H. (2018). Structure and Conformational Dynamics of the Human Spliceosomal Bact Complex. Cell *172*, 454–464.e11.

Hauri, S., Khakzad, H., Happonen, L., Teleman, J., Malmström, J., and Malmström, L. (2019). Rapid determination of quaternary protein structures in complex biological samples. Nat. Commun. *10*, 192.

Hennig, J., de Vries, S.J., Hennig, K.D., Randles, L., Walters, K.J., Sunnerhagen, M., and Bonvin, A.M.J.J. (2012). MTMDAT-HADDOCK: high-throughput, protein complex structure modeling based on limited proteolysis and mass spectrometry. BMC Struct. Biol. *12*, 29.

Herzog, F., Kahraman, A., Boehringer, D., Mak, R., Bracher, A., Walzthoeni, T., Leitner, A., Beck, M., Hartl, F.-U., Ban, N., et al. (2012). Structural probing of a protein phosphatase 2A network by chemical cross-linking and mass spectrometry. Science *337*, 1348–1352.

Hevler, J.F., Lukassen, M.V., Cabrera-Orefice, A., Arnold, S., Pronker, M.F., Franc, V., and Heck, A.J.R. (2021). Selective cross-linking of coinciding protein assemblies by in-gel cross-

linking mass spectrometry. EMBO J. *40*, e106174.

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

Huang, B.X., and Kim, H.-Y. (2009). Probing Akt-inhibitor interaction by chemical cross-linking and mass spectrometry. J. Am. Soc. Mass Spectrom. *20*, 1504–1513.

Humphreys, I.R., Pei, J., Baek, M., Krishnakumar, A., Anishchenko, I., Ovchinnikov, S., Zhang, J., Ness, T.J., Banjade, S., Bagde, S., et al. (2021). Structures of core eukaryotic protein complexes.

Iacobucci, C., and Sinz, A. (2017). To Be or Not to Be? Five Guidelines to Avoid Misassignments in Cross-Linking/Mass Spectrometry. Anal. Chem. *89*, 7832–7835.

Iacobucci, C., Götze, M., Ihling, C.H., Piotrowski, C., Arlt, C., Schäfer, M., Hage, C., Schmidt, R., and Sinz, A. (2018). A cross-linking/mass spectrometry workflow based on MS-cleavable cross-linkers and the MeroX software for studying protein structures and protein–protein interactions. Nat. Protoc. *13*, 2864–2889.

James, G.T. (1978). Inactivation of the protease inhibitor phenylmethylsulfonyl fluoride in buffers. Anal. Biochem. *86*, 574–579.

Jishage, M., Yu, X., Shi, Y., Ganesan, S.J., Chen, W.-Y., Sali, A., Chait, B.T., Asturias, F.J., and Roeder, R.G. (2018). Architecture of Pol II (G) and molecular mechanism of transcription regulation by Gdown1. Nat. Struct. Mol. Biol. *25*, 859–867.

de Jong, L., Roseboom, W., and Kramer, G. (2021). A composite filter for low FDR of protein-protein interactions detected by in vivo cross-linking. J. Proteomics *230*, 103987.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature.

Kaake, R.M., Wang, X., Burke, A., Yu, C., Kandur, W., Yang, Y., Novtisky, E.J., Second, T., Duan, J., Kao, A., et al. (2014). A new in vivo cross-linking mass spectrometry platform to define protein-protein interactions in living cells. Mol. Cell. Proteomics *13*, 3533–3543.

Kahraman, A., Malmström, L., and Aebersold, R. (2011). Xwalk: computing and visualizing distances in cross-linking experiments. Bioinformatics *27*, 2163–2164.

Kahraman, A., Herzog, F., Leitner, A., Rosenberger, G., Aebersold, R., and Malmström, L. (2013). Cross-link guided molecular modeling with ROSETTA. PLoS One *8*, e73411.

Kalinin, S., Peulen, T., Sindbert, S., Rothwell, P.J., Berger, S., Restle, T., Goody, R.S., Gohlke, H., and Seidel, C.A.M. (2012). A toolkit and benchmark study for FRET-restrained high-precision structural modeling. Nature Methods *9*, 1218–1225.

Kao, A., Chiu, C.-L., Vellucci, D., Yang, Y., Patel, V.R., Guan, S., Randall, A., Baldi, P., Rychnovsky, S.D., and Huang, L. (2011). Development of a novel cross-linking strategy for fast and accurate identification of cross-linked peptides of protein complexes. Mol. Cell. Proteomics *10*, M110.002212.

Karaca, E., and Bonvin, A.M.J.J. (2013). Advances in integrative modeling of biomolecular complexes. Methods *59*, 372–381.

Karaca, E., Rodrigues, J.P.G.L.M., Graziadei, A., Bonvin, A.M.J.J., and Carlomagno, T. (2017). M3: an integrative framework for structure determination of molecular machines. Nat. Methods *14*, 897–902.

Karr, J.R., Sanghvi, J.C., Macklin, D.N., Gutschow, M.V., Jacobs, J.M., Bolival, B., Jr, Assad-Garcia, N., Glass, J.I., and Covert, M.W. (2012). A whole-cell computational model predicts phenotype from genotype. Cell *150*, 389–401.

Kastritis, P.L., O'Reilly, F.J., Bock, T., Li, Y., Rogon, M.Z., Buczak, K., Romanov, N., Betts, M.J., Bui, K.H., Hagen, W.J., et al. (2017). Capturing protein communities by structural proteomics in a thermophilic eukaryote. Molecular Systems Biology *13*, 936.

Kawabata, T. (2008). Multiple Subunit Fitting into a Low-Resolution Density Map of a Macromolecular Complex Using a Gaussian Mixture Model. Biophysical Journal *95*, 4643–4658.

Keller, A., Chavez, J.D., Felt, K.C., and Bruce, J.E. (2019). Prediction of an Upper Limit for the Fraction of Interprotein Cross-Links in Large-Scale In Vivo Cross-Linking Studies. J. Proteome Res. *18*, 3077–3085.

Kim, S.J., Fernandez-Martinez, J., Sampathkumar, P., Martel, A., Matsui, T., Tsuruta, H., Weiss, T.M., Shi, Y., Markina-Inarrairaegui, A., Bonanno, J.B., et al. (2014). Integrative structure-function mapping of the nucleoporin Nup133 suggests a conserved mechanism for membrane anchoring of the nuclear pore complex. Mol. Cell. Proteomics *13*, 2911–2926.

Kim, S.J., Fernandez-Martinez, J., Nudelman, I., Shi, Y., Zhang, W., Raveh, B., Herricks, T., Slaughter, B.D., Hogan, J.A., Upla, P., et al. (2018). Integrative structure and functional anatomy of a nuclear pore complex. Nature *555*, 475–482.

Kolbowski, L., Mendes, M.L., and Rappsilber, J. (2017). Optimizing the Parameters Governing the Fragmentation of Cross-Linked Peptides in a Tribrid Mass Spectrometer. Analytical Chemistry *89*, 5311–5318.

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

Kosinski, J., Mosalaganti, S., von Appen, A., Teimer, R., DiGuilio, A.L., Wan, W., Bui, K.H., Hagen, W.J.H., Briggs, J.A.G., Glavy, J.S., et al. (2016). Molecular architecture of the inner ring scaffold of the human nuclear pore complex. Science *352*, 363–365.

Koukos, P.I., and Bonvin, A.M.J.J. (2020). Integrative Modelling of Biomolecular Complexes. J. Mol. Biol. *432*, 2861–2881.

Kramer, K., Sachsenberg, T., Beckmann, B.M., Qamar, S., Boon, K.-L., Hentze, M.W., Kohlbacher, O., and Urlaub, H. (2014). Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins. Nat. Methods *11*, 1064–1070.

Kryshtafovych, A., Schwede, T., Topf, M., Fidelis, K., and Moult, J. (2019). Critical assessment of methods of protein structure prediction (CASP)—Round XIII. Proteins: Structure, Function, and Bioinformatics *87*, 1011–1020.

Kühlbrandt, W. (2014). Biochemistry. The resolution revolution. Science *343*, 1443–1444.

Kukacka, Z., Rosulek, M., Strohalm, M., Kavan, D., and Novak, P. (2015). Mapping protein

structural changes by quantitative cross-linking. Methods *89*, 112–120.

Lasker, K., Förster, F., Bohn, S., Walzthoeni, T., Villa, E., Unverdorben, P., Beck, F., Aebersold, R., Sali, A., and Baumeister, W. (2012). Molecular architecture of the 26S proteasome holocomplex determined by an integrative approach. Proc. Natl. Acad. Sci. U. S. A. *109*, 1380–1387.

Leaver-Fay, A., Tyka, M., Lewis, S.M., Lange, O.F., Thompson, J., Jacak, R., Kaufman, K.W., Renfrew, P.D., Smith, C.A., Sheffler, W., et al. (2011). ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. Methods Enzymol. *487*, 545–574.

Lee, B.-G., Merkel, F., Allegretti, M., Hassler, M., Cawood, C., Lecomte, L., O'Reilly, F.J., Sinn, L.R., Gutierrez-Escribano, P., Kschonsak, M., et al. (2020). Cryo-EM structures of holo condensin reveal a subunit flip-flop mechanism. Nat. Struct. Mol. Biol. *27*, 743–751.

Leitner, A., Reischl, R., Walzthoeni, T., Herzog, F., Bohn, S., Förster, F., and Aebersold, R. (2012). Expanding the chemical cross-linking toolbox by the use of multiple proteases and enrichment by size exclusion chromatography. Mol. Cell. Proteomics *11*, M111.014126.

Leitner, A., Walzthoeni, T., and Aebersold, R. (2014). 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.

Leitner, A., Faini, M., Stengel, F., and Aebersold, R. (2016). Crosslinking and Mass Spectrometry: An Integrated Technology to Understand the Structure and Function of Molecular Machines. Trends Biochem. Sci. *41*, 20–32.

Leitner, A., Bonvin, A.M.J.J., Borchers, C.H., Chalkley, R.J., Chamot-Rooke, J., Combe, C.W., Cox, J., Dong, M.-Q., Fischer, L., Götze, M., et al. (2020). Toward Increased Reliability, Transparency, and Accessibility in Cross-linking Mass Spectrometry. Structure *28*, 1259–1268.

Leman, J.K., Weitzner, B.D., Lewis, S.M., Adolf-Bryfogle, J., Alam, N., Alford, R.F., Aprahamian, M., Baker, D., Barlow, K.A., Barth, P., et al. (2020). Macromolecular modeling and design in Rosetta: recent methods and frameworks. Nat. Methods *17*, 665–680.

Lenz, S., Sinn, L.R., O'Reilly, F.J., Fischer, L., Wegner, F., and Rappsilber, J. (2021). Reliable identification of protein-protein interactions by crosslinking mass spectrometry. Nat. Commun. *12*, 3564.

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

Linden, A., Deckers, M., Parfentev, I., Pflanz, R., Homberg, B., Neumann, P., Ficner, R., Rehling, P., and Urlaub, H. (2020). A Cross-linking Mass Spectrometry Approach Defines Protein Interactions in Yeast Mitochondria. Mol. Cell. Proteomics *19*, 1161–1178.

Liu, F., Rijkers, D.T.S., Post, H., and Heck, A.J.R. (2015). Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry. Nat. Methods *12*, 1179–1184.

Liu, F., Lössl, P., Scheltema, R., Viner, R., and Heck, A.J.R. (2017a). Optimized fragmentation schemes and data analysis strategies for proteome-wide cross-link identification. Nature Communications *8*.

Liu, F., Lössl, P., Rabbitts, B.M., Balaban, R.S., and Heck, A.J.R. (2018). The interactome of intact mitochondria by cross-linking mass spectrometry provides evidence for coexisting respiratory supercomplexes. Mol. Cell. Proteomics *17*, 216–232.

Liu, Q., Remmelzwaal, S., Heck, A.J.R., Akhmanova, A., and Liu, F. (2017b). Facilitating identification of minimal protein binding domains by cross-linking mass spectrometry. Sci. Rep. *7*, 13453.

Lively, M.O., and Powers, J.C. (1978). Specificity and reactivity of human granulocyte elastase and cathepsin G, porcine pancreatic elastase, bovine chymotrypsin and trypsin toward inhibition with sulfonyl flourides. Biochimica et Biophysica Acta (BBA) - Enzymology *525*, 171–179.

Low, T.Y., Syafruddin, S.E., Mohtar, M.A., Vellaichamy, A., A Rahman, N.S., Pung, Y.-F., and Tan, C.S.H. (2021). Recent progress in mass spectrometry-based strategies for elucidating protein–protein interactions. Cell. Mol. Life Sci. *78*, 5325–5339.

Luchinat, E., and Banci, L. (2017). In-cell NMR: a topical review. IUCrJ *4*, 108–118.

Mädler, S., Bich, C., Touboul, D., and Zenobi, R. (2009). Chemical cross-linking with NHS esters: a systematic study on amino acid reactivities. J. Mass Spectrom. *44*, 694–706.

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

Makowski, M.M., Willems, E., Jansen, P.W.T.C., and Vermeulen, M. (2016). Cross-linking immunoprecipitation-MS (xIP-MS): Topological Analysis of Chromatin-associated Protein Complexes Using Single Affinity Purification. Mol. Cell. Proteomics *15*, 854–865.

Marze, N.A., Roy Burman, S.S., Sheffler, W., and Gray, J.J. (2018). Efficient flexible backbone protein–protein docking for challenging targets. Bioinformatics *34*, 3461–3469.

Masrati, G., Landau, M., Ben-Tal, N., Lupas, A., Kosloff, M., and Kosinski, J. (2021). Integrative Structural Biology in the Era of Accurate Structure Prediction. J. Mol. Biol. 167127.

Mateus, A., Kurzawa, N., Becher, I., Sridharan, S., Helm, D., Stein, F., Typas, A., and Savitski, M.M. (2020). Thermal proteome profiling for interrogating protein interactions. Mol. Syst. Biol. *16*, e9232.

Matzinger, M., and Mechtler, K. (2021). Cleavable Cross-Linkers and Mass Spectrometry for the Ultimate Task of Profiling Protein–Protein Interaction Networks in Vivo. J. Proteome Res. *20*, 78–93.

Mendes, M.L., Fischer, L., Chen, Z.A., Barbon, M., O'Reilly, F.J., Giese, S.H., Bohlke-Schneider, M., Belsom, A., Dau, T., Combe, C.W., et al. (2019). An integrated workflow for crosslinking mass spectrometry. Mol. Syst. Biol. *15*, e8994.

Merkley, E.D., Rysavy, S., Kahraman, A., Hafen, R.P., Daggett, V., and Adkins, J.N. (2014). Distance restraints from crosslinking mass spectrometry: mining a molecular dynamics simulation database to evaluate lysine-lysine distances. Protein Sci. *23*, 747–759.

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

Mirdita, M., Ovchinnikov, S., and Steinegger, M. ColabFold - Making protein folding

accessible to all.

Molnar, K.S., Bonomi, M., Pellarin, R., Clinthorne, G.D., Gonzalez, G., Goldberg, S.D., Goulian, M., Sali, A., and DeGrado, W.F. (2014). Cys-scanning disulfide crosslinking and bayesian modeling probe the transmembrane signaling mechanism of the histidine kinase, PhoQ. Structure *22*, 1239–1251.

Müller, M.Q., Dreiocker, F., Ihling, C.H., Schäfer, M., and Sinz, A. (2010). Cleavable cross-linker for protein structure analysis: reliable identification of cross-linking products by tandem MS. Anal. Chem. *82*, 6958–6968.

Narayanan, A., and Jones, L.H. (2015). Sulfonyl fluorides as privileged warheads in chemical biology. Chem. Sci. *6*, 2650–2659.

Niemeyer, M., Moreno Castillo, E., Ihling, C.H., Iacobucci, C., Wilde, V., Hellmuth, A., Hoehenwarter, W., Samodelov, S.L., Zurbriggen, M.D., Kastritis, P.L., et al. (2020). Flexibility of intrinsically disordered degrons in AUX/IAA proteins reinforces auxin co-receptor assemblies. Nat. Commun. *11*, 2277.

Orbán-Németh, Z., Beveridge, R., Hollenstein, D.M., Rampler, E., Stranzl, T., Hudecz, O., Doblmann, J., Schlögelhofer, P., and Mechtler, K. (2018). Author Correction: Structural prediction of protein models using distance restraints derived from cross-linking mass spectrometry data. Nat. Protoc. *13*, 1724.

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

O'Reilly, F.J., Xue, L., Graziadei, A., Sinn, L., Lenz, S., Tegunov, D., Blötz, C., Singh, N., Hagen, W.J.H., Cramer, P., et al. (2020). In-cell architecture of an actively transcribing-translating expressome. Science *369*, 554–557.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H., and Ferrin, T.E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci. *30*, 70–82.

Piersimoni, L., and Sinz, A. (2020). Cross-linking/mass spectrometry at the crossroads. Anal. Bioanal. Chem. *412*, 5981–5987.

Piotrowski, C., and Sinz, A. (2018). Structural Investigation of Proteins and Protein Complexes by Chemical Cross-Linking/Mass Spectrometry. Advances in Experimental Medicine and Biology 101–121.

Politis, A., Schmidt, C., Tjioe, E., Sandercock, A.M., Lasker, K., Gordiyenko, Y., Russel, D., Sali, A., and Robinson, C.V. (2015). Topological models of heteromeric protein assemblies from mass spectrometry: application to the yeast eIF3:eIF5 complex. Chem. Biol. *22*, 117–128.

Rappsilber, J. (2011). The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. J. Struct. Biol. *173*, 530–540.

Raveh, B., Sun, L., White, K.L., Sanyal, T., Tempkin, J., Zheng, D., Bharath, K., Singla, J., Wang, C., Zhao, J., et al. (2021). Bayesian metamodeling of complex biological systems across varying representations. Proc. Natl. Acad. Sci. U. S. A. *118*.

Rey, M., Dhenin, J., Kong, Y., Nouchikian, L., Filella, I., Duchateau, M., Dupré, M., Pellarin, R., Duménil, G., and Chamot-Rooke, J. (2021). Advanced In Vivo Cross-Linking Mass Spectrometry Platform to Characterize Proteome-Wide Protein Interactions. Analytical Chemistry *93*, 4166–4174.

Rieping, W., Habeck, M., and Nilges, M. (2005). Inferential structure determination. Science *309*, 303–306.

Rinner, O., Seebacher, J., Walzthoeni, T., Mueller, L.N., Beck, M., Schmidt, A., Mueller, M., and Aebersold, R. (2008). Identification of cross-linked peptides from large sequence databases. Nat. Methods *5*, 315–318.

Ritorto, M.S., Cook, K., Tyagi, K., Pedrioli, P.G.A., and Trost, M. (2013). Hydrophilic strong anion exchange (hSAX) chromatography for highly orthogonal peptide separation of complex proteomes. J. Proteome Res. *12*, 2449–2457.

Rodrigues, J.P.G.L.M., Trellet, M., Schmitz, C., Kastritis, P., Karaca, E., Melquiond, A.S.J., and Bonvin, A.M.J.J. (2012). Clustering biomolecular complexes by residue contacts similarity. Proteins *80*, 1810–1817.

Roel-Touris, J., Don, C.G., Honorato, R., V., Rodrigues, J.P.G.L.M., and Bonvin, A.M.J.J. (2019). Less Is More: Coarse-Grained Integrative Modeling of Large Biomolecular Assemblies with HADDOCK. J. Chem. Theory Comput. *15*, 6358–6367.

Rout, M.P., and Sali, A. (2019). Principles for Integrative Structural Biology Studies. Cell *177*, 1384–1403.

Russel, D., Lasker, K., Webb, B., Velázquez-Muriel, J., Tjioe, E., Schneidman-Duhovny, D., Peterson, B., and Sali, A. (2012). Putting the pieces together: integrative modeling platform software for structure determination of macromolecular assemblies. PLoS Biol. *10*, e1001244.

Ryl, P.S.J., Bohlke-Schneider, M., Lenz, S., Fischer, L., Budzinski, L., Stuiver, M., Mendes, M.M.L., Sinn, L., O'Reilly, F.J., and Rappsilber, J. (2020). In Situ Structural Restraints from Cross-Linking Mass Spectrometry in Human Mitochondria. J. Proteome Res. *19*, 327–336.

Saltzberg, D.J., Broughton, H.B., Pellarin, R., Chalmers, M.J., Espada, A., Dodge, J.A., Pascal, B.D., Griffin, P.R., Humblet, C., and Sali, A. (2017). A Residue-Resolved Bayesian Approach to Quantitative Interpretation of Hydrogen-Deuterium Exchange from Mass Spectrometry: Application to Characterizing Protein-Ligand Interactions. J. Phys. Chem. B *121*, 3493–3501.

Schiffrin, B., Radford, S.E., Brockwell, D.J., and Calabrese, A.N. (2020). PyXlinkViewer: A flexible tool for visualization of protein chemical crosslinking data within the PyMOL molecular graphics system. Protein Sci. *29*, 1851–1857.

Schmidt, C., and Robinson, C.V. (2014). A comparative cross-linking strategy to probe conformational changes in protein complexes. Nat. Protoc. *9*, 2224–2236.

Schmidt, C., and Urlaub, H. (2017). Combining cryo-electron microscopy (cryo-EM) and cross-linking mass spectrometry (CX-MS) for structural elucidation of large protein assemblies. Curr. Opin. Struct. Biol. *46*, 157–168.

Schneidman-Duhovny, D., Pellarin, R., and Sali, A. (2014). Uncertainty in integrative structural modeling. Curr. Opin. Struct. Biol. *28*, 96–104.

Schnirch, L., Nadler-Holly, M., Siao, S.-W., Frese, C.K., Viner, R., and Liu, F. (2020). Expanding the Depth and Sensitivity of Cross-Link Identification by Differential Ion Mobility Using High-Field Asymmetric Waveform Ion Mobility Spectrometry. Anal. Chem. *92*, 10495–10503.

Schopper, S., Kahraman, A., Leuenberger, P., Feng, Y., Piazza, I., Müller, O., Boersema, P.J., and Picotti, P. (2017). Measuring protein structural changes on a proteome-wide scale using limited proteolysis-coupled mass spectrometry. Nat. Protoc. *12*, 2391–2410.

Schur, F.K. (2019). Toward high-resolution in situ structural biology with cryo-electron tomography and subtomogram averaging. Curr. Opin. Struct. Biol. *58*, 1–9.

Schweppe, D.K., Zheng, C., Chavez, J.D., Navare, A.T., Wu, X., Eng, J.K., and Bruce, J.E. (2016). XLinkDB 2.0: integrated, large-scale structural analysis of protein crosslinking data. Bioinformatics *32*, 2716–2718.

Schweppe, D.K., Chavez, J.D., Lee, C.F., Caudal, A., Kruse, S.E., Stuppard, R., Marcinek, D.J., Shadel, G.S., Tian, R., and Bruce, J.E. (2017). Mitochondrial protein interactome elucidated by chemical cross-linking mass spectrometry. Proc. Natl. Acad. Sci. U. S. A. *114*, 1732–1737.

Shi, Y., Fernandez-Martinez, J., Tjioe, E., Pellarin, R., Kim, S.J., Williams, R., Schneidman-Duhovny, D., Sali, A., Rout, M.P., and Chait, B.T. (2014). Structural characterization by cross-linking reveals the detailed architecture of a coatomer-related heptameric module from the nuclear pore complex. Mol. Cell. Proteomics *13*, 2927–2943.

Shin, M., Puchades, C., Asmita, A., Puri, N., Adjei, E., Wiseman, R.L., Karzai, A.W., and Lander, G.C. (2020). Structural basis for distinct operational modes and protease activation in AAA+ protease Lon. Sci Adv *6*, eaba8404.

Singh, J., Ponnaiyan, S., Gieselmann, V., and Winter, D. (2021). Generation of Antibodies Targeting Cleavable Cross-Linkers. Anal. Chem. *93*, 3762–3769.

Singla, J., McClary, K.M., White, K.L., Alber, F., Sali, A., and Stevens, R.C. (2018). Opportunities and Challenges in Building a Spatiotemporal Multi-scale Model of the Human Pancreatic β Cell. Cell *173*, 11–19.

Sinnott, M., Malhotra, S., Madhusudhan, M.S., Thalassinos, K., and Topf, M. (2020). Combining Information from Crosslinks and Monolinks in the Modeling of Protein Structures. Structure *28*, 1061–1070.e3.

Sinz, A. (2018). Crosslinking Mass Spectrometry Goes In-Tissue. Cell Syst *6*, 10–12.

Skinnider, M.A., and Foster, L.J. (2021). Meta-analysis defines principles for the design and analysis of co-fractionation mass spectrometry experiments. Nat. Methods *18*, 806–815.

Slavin, M., Zamel, J., Zohar, K., Eliyahu, T., Braitbard, M., Brielle, E., Baraz, L., Stolovich-Rain, M., Friedman, A., Wolf, D.G., et al. (2021). Targeted in situ cross-linking mass spectrometry and integrative modeling reveal the architectures of three proteins from SARS-CoV-2. Proc. Natl. Acad. Sci. U. S. A. *118*.

Stadlmeier, M., Runtsch, L.S., Streshnev, F., Wühr, M., and Carell, T. (2020). A Click-Chemistry-Based Enrichable Crosslinker for Structural and Protein Interaction Analysis by

Mass Spectrometry. Chembiochem *21*, 103–107.

Steigenberger, B., Pieters, R.J., Heck, A.J.R., and Scheltema, R.A. (2019). PhoX: An IMAC-Enrichable Cross-Linking Reagent. ACS Cent Sci *5*, 1514–1522.

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

Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P., et al. (2018). STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. *47*, D607–D613.

Tang, X., and Bruce, J.E. (2010). A new cross-linking strategy: protein interaction reporter (PIR) technology for protein–protein interaction studies. Molecular BioSystems *6*, 939.

Tang, X., Wippel, H.H., Chavez, J.D., and Bruce, J.E. (2021). Crosslinking mass spectrometry: A link between structural biology and systems biology. Protein Sci. *30*, 773–784.

Townsend, C., Leelaram, M.N., Agafonov, D.E., Dybkov, O., Will, C.L., Bertram, K., Urlaub, H., Kastner, B., Stark, H., and Lührmann, R. (2020). Mechanism of protein-guided folding of the active site U2/U6 RNA during spliceosome activation. Science *370*.

Trester-Zedlitz, M., Kamada, K., Burley, S.K., Fenyö, D., Chait, B.T., and Muir, T.W. (2003). A modular cross-linking approach for exploring protein interactions. J. Am. Chem. Soc. *125*, 2416–2425.

Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Žídek, A., Bridgland, A., Cowie, A., Meyer, C., Laydon, A., et al. (2021). Highly accurate protein structure prediction for the human proteome. Nature.

Tüting, C., Iacobucci, C., Ihling, C.H., Kastritis, P.L., and Sinz, A. (2020). Structural analysis of 70S ribosomes by cross-linking/mass spectrometry reveals conformational plasticity. Sci. Rep. *10*.

Vallat, B., Webb, B., Westbrook, J., Sali, A., and Berman, H.M. (2019). Archiving and disseminating integrative structure models. Journal of Biomolecular NMR *73*, 385–398.

Viswanath, S., Chemmama, I.E., Cimermancic, P., and Sali, A. (2017a). Assessing Exhaustiveness of Stochastic Sampling for Integrative Modeling of Macromolecular Structures. Biophys. J. *113*, 2344–2353.

Viswanath, S., Bonomi, M., Kim, S.J., Klenchin, V.A., Taylor, K.C., Yabut, K.C., Umbreit, N.T., Van Epps, H.A., Meehl, J., Jones, M.H., et al. (2017b). The molecular architecture of the yeast spindle pole body core determined by Bayesian integrative modeling. Mol. Biol. Cell *28*, 3298–3314.

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

Walzthoeni, T., Joachimiak, L.A., Rosenberger, G., Röst, H.L., Malmström, L., Leitner, A., Frydman, J., and Aebersold, R. (2015). xTract: software for characterizing conformational changes of protein complexes by quantitative cross-linking mass spectrometry. Nat. Methods *12*, 1185–1190.

Wang, J., Burdzinski, G., Kubicki, J., Platz, M.S., Moss, R.A., Fu, X., Piotrowiak, P., and Myahkostupov, M. (2006). Ultrafast spectroscopic study of the photochemistry and photophysics of arylhalodiazirines: direct observation of carbene and zwitterion formation. J. Am. Chem. Soc. *128*, 16446–16447.

Ward, A.B., Sali, A., and Wilson, I.A. (2013). Integrative Structural Biology. Science *339*, 913–915.

West, A.V., Muncipinto, G., Wu, H.-Y., Huang, A.C., Labenski, M.T., Jones, L.H., and Woo, C.M. (2021). Labeling Preferences of Diazirines with Protein Biomolecules. J. Am. Chem. Soc. *143*, 6691–6700.

Wittig, S., Ganzella, M., Barth, M., Kostmann, S., Riedel, D., Pérez-Lara, Á., Jahn, R., and Schmidt, C. (2021). Cross-linking mass spectrometry uncovers protein interactions and functional assemblies in synaptic vesicle membranes. Nat. Commun. *12*, 1–14.

Woetzel, N., Lindert, S., Stewart, P.L., and Meiler, J. (2011). BCL::EM-Fit: Rigid body fitting of atomic structures into density maps using geometric hashing and real space refinement. Journal of Structural Biology *175*, 264–276.

Wu, X., Chavez, J.D., Schweppe, D.K., Zheng, C., Weisbrod, C.R., Eng, J.K., Murali, A., Lee, S.A., Ramage, E., Gallagher, L.A., et al. (2016). In vivo protein interaction network analysis reveals porin-localized antibiotic inactivation in Acinetobacter baumannii strain AB5075. Nat. Commun. *7*, 13414.

Xie, Y., Clarke, B.P., Kim, Y.J., Ivey, A.L., Hill, P.S., Shi, Y., and Ren, Y. (2021). Cryo-EM structure of the yeast TREX complex and coordination with the SR-like protein Gbp2. Elife *10*.

Yan, K., Yang, J., Zhang, Z., McLaughlin, S.H., Chang, L., Fasci, D., Ehrenhofer-Murray, A.E., Heck, A.J.R., and Barford, D. (2019). Structure of the inner kinetochore CCAN complex assembled onto a centromeric nucleosome. Nature *574*, 278–282.

Yang, B., Wu, H., Schnier, P.D., Liu, Y., Liu, J., Wang, N., DeGrado, W.F., and Wang, L. (2018). Proximity-enhanced SuFEx chemical cross-linker for specific and multitargeting cross-linking mass spectrometry. Proc. Natl. Acad. Sci. U. S. A. *115*, 11162–11167.

Yang, J., Anishchenko, I., Park, H., Peng, Z., Ovchinnikov, S., and Baker, D. (2020). Improved protein structure prediction using predicted interresidue orientations. Proc. Natl. Acad. Sci. U. S. A. *117*, 1496–1503.

Yu, C., and Huang, L. (2018). Cross-Linking Mass Spectrometry: An Emerging Technology for Interactomics and Structural Biology. Anal. Chem. *90*, 144–165.

Yu, C., Huszagh, A., Viner, R., Novitsky, E.J., Rychnovsky, S.D., and Huang, L. (2016). Developing a Multiplexed Quantitative Cross-Linking Mass Spectrometry Platform for Comparative Structural Analysis of Protein Complexes. Anal. Chem. *88*, 10301–10308.

Yu, C., Novitsky, E.J., Cheng, N.W., Rychnovsky, S.D., and Huang, L. (2020). Exploring Spacer Arm Structures for Designs of Asymmetric Sulfoxide-Containing MS-Cleavable Cross-Linkers. Anal. Chem. *92*, 6026–6033.

Yugandhar, K., Wang, T.-Y., Wierbowski, S.D., Shayhidin, E.E., and Yu, H. (2020). Structure-based validation can drastically underestimate error rate in proteome-wide cross-linking mass spectrometry studies. Nat. Methods *17*, 985–988.

Zheng, Q., Zhang, H., Wu, S., and Chen, H. (2016). Probing Protein 3D Structures and Conformational Changes Using Electrochemistry-Assisted Isotope Labeling Cross-Linking Mass Spectrometry. J. Am. Soc. Mass Spectrom. *27*, 864–875.

Ziemianowicz, D.S., Ng, D., Schryvers, A.B., and Schriemer, D.C. (2019). Photo-Cross-Linking Mass Spectrometry and Integrative Modeling Enables Rapid Screening of Antigen Interactions Involving Bacterial Transferrin Receptors. J. Proteome Res. *18*, 934–946.

van Zundert, G.C.P., and Bonvin, A.M.J.J. (2015). DisVis: quantifying and visualizing accessible interaction space of distance-restrained biomolecular complexes. Bioinformatics *31*, 3222–3224.

van Zundert, G.C.P., Melquiond, A.S.J., and Bonvin, A.M.J.J. (2015). Integrative Modeling of Biomolecular Complexes: HADDOCKing with Cryo-Electron Microscopy Data. Structure *23*, 949–960.

van Zundert, G.C.P., Trellet, M., Schaarschmidt, J., Kurkcuoglu, Z., David, M., Verlato, M., Rosato, A., and Bonvin, A.M.J.J. (2017). The DisVis and PowerFit Web Servers: Explorative and Integrative Modeling of Biomolecular Complexes. J. Mol. Biol. *429*, 399–407.