Published in final edited form as:
J Struct Biol. 2011 March ; 173(3): 530–540. doi:10.1016/j.jsb.2010.10.014.

The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes

Juri Rappsilber
Wellcome Trust Centre for Cell Biology, University of Edinburgh, Michael Swann Building, King's Buildings, Mayfield Road, Edinburgh EH9 3JR, SCOTLAND, UK

Abstract

After more than a decade of method development, cross-linking in combination with mass spectrometry and bioinformatics is finally coming of age. This technology now provides improved opportunities for modelling by mapping structural details of functional complexes in solution. The structure of proteins or protein complexes is ascertained by identifying amino acid pairs that are positioned in close proximity to each other. The validity of this technique has recently been benchmarked for large multi-protein complexes, by comparing cross-link data with that from a crystal structure of RNA polymerase II. Here, the specific nature of this cross-linking data will be discussed to assess the technical challenges and opportunities for model building. We believe that once cross-linking/mass spectrometry data has been incorporated into modelling algorithms it will quickly become an indispensable companion of protein and protein complex modelling and a corner-stone of integrated structural biology.

Keywords

Cross-linking; mass spectrometry; modelling; integrated structural biology; multi-protein complexes

1. Introduction

1.1. Background

Cross-linking converts non-covalent interactions between proteins or simply their proximity into covalent bonds. The artificially fused molecules withstand denaturating conditions and thus can be analysed using methods that normally dissociate protein complexes. As early as in the 1970s, this revealed protein-protein contacts in ribosomes through the pairing of cross-linking with gel electrophoretic approaches (Clegg and Hayes, 1974; Sun et al., 1974). Nearly 30 years later, the arrival of peptide mass spectrometry (MS) and its transforming powers on all fields of life sciences (Aebersold and Mann, 2003) provided the impetus to develop cross-linking methods (reviewed by (Back et al., 2003; Trakselis et al., 2005; Sinz, 2006; Sinz, 2007; Mouradov et al., 2008; Tang and Bruce, 2009; Singh et al., 2010)). MS promised to efficiently identify the cross-linked proteins and furthermore to reveal precisely which residues were involved in the cross-link. In order to be cross-linked residues must be within a certain distance of each other, (determined by the cross-linking agent used), and this therefore provides a valuable experimental constraint for any modelling attempt.

juri.rappsilber@ed.ac.uk Phone: +44-(0)131-651-7056 FAX: +44-(0)131-650-5379.

Cross-linking and MS were used to provide a topological map of the Nup84 complex by gel electrophoretically separating and identifying cross-linked proteins (Rappsilber et al., 2000). This method has since been confirmed by the crystallographically characterized yeast 20S proteasome core (Denison and Kodadek, 2004) and been used for the analysis of the 19S proteasome lid (Sharon et al., 2006). These studies have shown the method to be a fast and reliable tool of proteomics, relying on protein identification as an established technology. A similar level of protein pair-wise interactions can also be obtained in a complementary way analysing native complexes by MS (Benesch and Robinson, 2006; Benesch et al., 2007; Heck, 2008).

Knowledge of the actual linkage sites, however, would increase the resolution of the method for structure determination from proteins to domains or even smaller sections, recently dubbed “peptide-level resolution” (Chen et al., 2010). Cross-linked amino acids were identified and used as distance constraints in conjunction with threading to determine the fold of a protein or protein domain (Young et al., 2000). Mass spectrometers, protocols, and algorithms have advanced since these first experiments a decade ago, such that cross-linking/MS can now be employed for the structural analysis of multi-protein complexes, even if their analysis proved challenging by other methods. Complexes that have been studied using these tools range from protein-peptide to large multi-protein complexes and include the Ffh.FtsY complex (Chu et al., 2004), GRP94, the endoplasmic reticulum Hsp90 (Chu et al., 2006b), the Ndc80 complex (Maiolica et al., 2007), the annexin A2/p11 complex (Schulz et al., 2007), an epitope-antibody complex (Pimenova et al., 2008), the calmodulin-Munc13 complex (Dimova et al., 2009), the phi29 connector/scaffolding complex (Fu et al., 2010), the GroEL-GroES chaperonin complex (Trnka and Burlingame, 2010), and RNA polymerase II alone and in complex with transcription factor IIF (Chen et al., 2010) (discussed in Section 3).

1.2. Current experience with cross-linking/mass spectrometry

Cross-linking/MS has a number of strengths (for challenges see Section 4). First and foremost, the analysis takes place in solution and focuses on large structures, i.e. provides data on proteins and domains in their native and quaternary structure. Heterogeneity in the sample as a result of multiple conformations, complex populations with differing subunit composition, or presence of other proteins may lengthen the analysis time and challenge the data interpretation but do not principally impair the study. In pioneering studies, proteins have been cross-linked in bacterial whole cell lysates (Rinner et al., 2008) and cell membranes of living bacteria (Zhang et al., 2009). The method is applicable to a wide selection of structural motifs including the otherwise difficult to study coiled coils (Maiolica et al., 2007) and possibly partially unfolded regions, although some folding appears to be required (Chen et al., 2010). Also, conformational changes in proteins have been studied in solution as compared to the crystal structure for the membrane protein rhodopsin (Jacobsen et al., 2006) or induced by binding of small molecules (Muller et al., 2009). Finally, cross-linking is fast and economical, while mass spectrometers are widely available for proteomic applications. Developers around the globe are tackling the current bottleneck of cross-linking/MS, namely the computational search tools for the identification of cross-linked peptides (see Section 2.4).

1.3. A potential link between cross-linking/mass spectrometry and modelling

Cross-linking/MS data have been used in conjunction with modelling for example to support homology modelling (Young et al., 2000; Chen et al., 2010) and to expand the crystal structure of the stable Pol II core towards the more dynamic periphery of a bound transcription factor, TFIIF (Chen et al., 2010). These and other individual applications (Fu et al., 2010) lack an automated framework. Nevertheless improvements in cross-linking/MS methods are expected to expedite many if not all aspects of modelling. First platforms to link cross-link data and modelling are being developed (Heymann et al., 2008). Model building based on x-ray diffraction may benefit when fitting protein chains into patchy regions of a density map or positioning un-observed protein regions such as loops, trimmed or truncated sequences, or missing sub-units. Docking of proteins may move from binary systems using e.g. HADDOCK (Karaca et al., 2010) to larger systems. Similarly, cross-link data may provide the intermediate resolution range currently lacking for the reconstruction of multi-domain proteins or multi-protein complexes from individual structure fragments obtained by x-ray crystallography, nuclear magnetic resonance (NMR), or homology modelling and out-line/shape revealing methods such as electron microscopic approaches or small angle x-ray scattering. In even larger assemblies, cross-linking may substitute for protein co-purification data such as used in reconstructing the nuclear pore complex (Alber et al., 2007a; Alber et al., 2007b). Last but not least, cross-linking opens a road towards dynamic aspects of proteins and multi-protein complexes. As an example, conformational changes could be modelled starting from a high-resolution structure of one conformation and a cross-linking/MS analysis of another conformation. This list will hopefully motivate developers of modelling tools to integrate cross-linking/MS data into the modelling process to reduce this current “bottleneck”.

Progression from early proof-of-concept experiments to the advent of routine application of cross-linking in structural biology requires a number of key challenges to be addressed. The experimental workflow of cross-linking/mass spectrometry will be outlined here. Discussion of the results of a recent detailed analysis of two large multi-protein complexes, Pol II and Pol II-TFIIF, will highlight practical details of the cross-link approach. Finally for researchers planning an experiment, interpreting results or using data for modelling a set of conclusions will be presented that summarize current knowledge on cross-link data.

2. Analytical workflow of cross-linking/mass spectrometry

The basic workflow to yield structural information of proteins and protein complexes by cross-linking/mass spectrometry (MS) is composed of four steps (Figure 1A). Proteins are cross-linked in solution and then digested by trypsin to give peptides, some of which will be cross-linked. This mixture of peptides is then analysed by mass spectrometry and resulting data is interpreted to identify cross-linked peptides and determine the linked residues.

2.1. Protein cross-linking

Proteins are cross-linked typically in a chemical reaction involving a cross-linker and side chains of amino acids. The reactivity of amino groups, thiols and carboxylic acids render them as prime targets for cross-linking. The cross-linker is typically a molecule with two reactive groups on either end separated by a spacer (Figure 1B). These reactive groups can either target amino groups (found in the side chain of lysine and at the protein N-terminus) (Figure 1C) or thiols (cysteine side chain). In published work to date, cross-linkers targeting exclusively amino groups have been used in cross-linking/MS studies of multi-protein complexes due to the high frequency of lysine in proteins and the consequently increased chance of obtaining and identifying cross-links. Alternatively, photo-activatable groups can be used in a cross-linker with currently poorly defined but presumably lower specificity (Krauth et al., 2009; Gomes and Gozzo, 2010). The result is always that the cross-linker bridges between residues within a protein or between two proteins at a maximal distance influenced by the length of the spacer. In a single exception, a small molecule, EDC, is used to activate carboxylic acids (aspartate, glutamate, protein C-terminus) to cross-link with amines (lysine, protein N-terminus). This directly cross-links atoms of the protein(s) with each other in a “zero-length” cross-link. Cross-linkers with three reactive groups exist but have not yet been used in structural work as they greatly increase the analytical challenges involved in identifying the three cross-linked amino acid residues. Cross-linkers are commercially available from several companies. New cross-linkers are being developed with improved chemical (Bich et al., 2010) or mass spectrometric properties (Petrotchenko et al., 2005; Tang et al., 2005; Chowdhury et al., 2006; Ihling et al., 2006; Gardner et al., 2008; Lu et al., 2008; Krauth et al., 2009; Paramelle et al., 2009; Petrotchenko et al., 2009; Dreiocker et al., 2010; Liu and Goshe, 2010; Petrotchenko et al., 2010; Yang et al., 2010; Zelter et al., 2010).

2.2. Digestion of cross-linked proteins to peptides

The identification of cross-link sites employs the well-established workflows of proteomics, but with a twist. Proteins are digested by trypsin into peptides which can be fractionated or separated but ultimately are analysed by mass spectrometry to determine their mass and usually also fragmentation spectra (Figure 1A). Standard proteomics analysis deals only with linear peptides in its efforts to identify and quantify proteins and to determine their modification sites. To these, cross-linking adds a number of different species (Figure 1D). At the protein level, cross-linking results in two products: a cross-link, when the cross-linker reacted with one amino acid on either end, or a modification, when the cross-linker reacted with an amino acid on one and water on the other end. At the peptide level, this can lead to three different situations and their combinations (Figure 1D): modified peptides (type 0, nomenclature by (Schilling et al., 2003)), cyclic or internally bridged peptides (type 1), cross-linked peptides (type 2), or any combination of these (type 3). All of these peptides contain structural information. The current focus is on cross-linked peptides (type 2) as they contain long distance information. In contrast, modified peptides (type 0) encode accessibility while cyclic peptides (type 1) encode local structure such as alpha-helical regions (Maiolica et al., 2007). Higher order cross-links (type 3) have yet to be observed and will likely be difficult to identify due to complex fragmentation spectra. Methods that distinguish during mass spectrometric detection between different cross-link products include isotope labelling schemes (Back et al., 2002; Chu et al., 2006a) and special cross-linker chemistry (Petrotchenko et al., 2005).

2.3. Mass spectrometric analysis of cross-linked peptides

MS provides the data to identify cross-linked residues in a two-staged process. First, the cross-linked peptide needs to be identified. For this, the mass and usually also the fragmentation spectrum of the cross-linked peptide have to be acquired and then analysed by database searching. Detailed analysis of the fragmentation spectrum may then reveal the exact or approximate sites of linkage, depending primarily on the quality and dynamic range of the spectrum. The analysis of peptide fragmentation spectra in general is simplified by the fact that peptides normally follow specific fragmentation rules, breaking predominantly along the backbone, at the peptide bond when using the most commonly employed fragmentation method, collision-induced dissociation (CID) (Wells and McLuckey, 2005). Peptide fragmentation by CID gives rise to two main fragment types, the N-terminal “b-ions” and the C-terminal “y-ions” (Roepstorff and Fohlman, 1984; Biemann, 1988). Peaks in fragmentation spectra are labelled using these letters, in conjunction with a subscript for the number of residues contained in the fragment and a superscript for the number and type (positive or negative) of charges of the ion. An alternative to CID is given by electron transfer dissociation (ETD) (Syka et al., 2004). In this case, c- and z-ions are observed predominantly. Note that mass spectrometry measures the mass to charge ratio of ions. The charge of an ion can be determined from resolved isotope peaks and the mass then be calculated (for more details on peptide fragmentation in a mass spectrometer consult the introductory review written by Steen and Mann (Steen and Mann, 2004)). Cross-linked peptides follow these general rules of peptide fragmentation by CID (Back et al., 2001; Schilling et al., 2003; Gaucher et al., 2006) and ETD (Chowdhury et al., 2009). The fragmentation spectrum of a cross-linked peptide typically features fragments of both cross-linked peptides (Figure 1E) and can thus lead to the confident and unambiguous identification of both peptides. If a set of fragments is observed that fall upstream and downstream of the cross-linked residues, the exact position of the cross-linking site can be determined. In the spectrum displayed in Figure 1E the fragments that determine the linkage sites are red b1 and y13 and green y4/y5 and b4/b5, respectively for the peptide sequences coloured correspondingly. Note that cross-linked peptides are best identified by a combination of high resolution for the peptide mass and high resolution for the fragment masses. This strategy, also called high-high, maximises the specificity of the database search. The identification of cross-linked peptides may furthermore be improved if specific reporter fragments are generated that are only observed in cross-linked peptides (Back et al., 2001; Seebacher et al., 2006; Iglesias et al., 2009; Iglesias et al., 2010) or new cross-linkers are used that guide the mass spectrometric analysis towards cross-linked peptides (Petrotchenko et al., 2005; Tang et al., 2005; Chowdhury et al., 2006; Ihling et al., 2006; Gardner et al., 2008; Lu et al., 2008; Krauth et al., 2009; Paramelle et al., 2009; Petrotchenko et al., 2009; Dreiocker et al., 2010; Liu and Goshe, 2010; Petrotchenko et al., 2010; Yang et al., 2010; Zelter et al., 2010).

For a long time, cross-linked peptides could not be, or when only poorly detected, despite their presence in the mixture of peptides obtained by trypsin digestion of cross-linked proteins or protein complexes. The multiple possible cross-link products for any specific residue and typically incomplete reaction results in low signals for cross-linked peptides. These need to be detected against a high background of unmodified linear peptides and possibly also non-specific reaction by-products. Three elements worked together to recently address the data acquisition challenge of cross-linked peptides: enrichment of cross-linked peptides, improved mass spectrometers, and automated data interpretation.

Enriching cross-linked peptides improves their detection by MS and thus the yield in observed linkage sites. Various methods to achieve such an enrichment have been envisaged and are now being tested. One such approach that has so far been employed for the analysis of multi-protein complexes makes use of the generally higher charge state that distinguishes cross-linked peptides from linear peptides. This has been exploited, prior to acquisition, by cation-exchange chromatography which enriches cross-linked peptides carrying higher charges in the later eluting fractions (Rinner et al., 2008; Chen et al., 2010) and during acquisition on the MS where peptides with high charge states are selected for fragmentation (Maiolica et al., 2007; Rinner et al., 2008; Chen et al., 2010). Numerous other approaches are currently under development, particularly the use of cross-linkers that contain affinity groups for the selective enrichment of cross-linked peptides (Chu et al., 2006a; Kasper et al., 2007; Chowdhury et al., 2009; Kang et al., 2009; Nessen et al., 2009; Yan et al., 2009; Vellucci et al., 2010).

A new generation of mass spectrometers has increased the number of peptide species that can be selected for fragmentation in a single experiment, the sensitivity of their detection and the resolution of signals. This results in more of the low-intensity cross-linked peptides being included in the analysis and in high-quality fragmentation data that can be interpreted unambiguously. To deal with all this date efficiently, computational approaches have been developed that automate the data interpretation step and thus allowed the power of liquid chromatography-coupled mass spectrometry (LC-MS) to be used to create large data sets for the detection of cross-linked peptides. This is the subject of the next section.

2.4. Identification of cross-linked peptides

Cross-linked peptides can be identified using mass spectrometry analogously to linear peptides. For linear peptides, the peptide mass is taken to select candidate peptides from a protein database matching this mass within the experimental error. The fragmentation spectrum is then used to find, from among these candidates, the peptide sequence that best explains the observed fragment signals. To adopt the same workflow to cross-linked peptides, all possible cross-linked peptides must be predicted by in-silico digestion of all proteins and then creation of all possible pair-wise combinations of peptides. Any peptide needs to be considered if it contains a residue that is capable of cross-linking in the actual experiment to be analysed. The pairing leads to $(n^2+n)/2$ possible cross-links for n peptides. This n^2 problem creates a challenge for the search algorithm and the evaluation of any match between a spectrum and a candidate peptide pair due to the danger of random matches in a large database. However, for protein complexes this problem is simplified, as only those proteins need to be considered that are actually present in the sample. The first automated algorithm that identified cross-links in a multi-protein complex (Maiolica et al., 2007) and the identification of cross-links in an E. coli cell lysate (Rinner et al., 2008) revealed no limitations to database searching of cross-linked peptides in principle. A large number of algorithms and programmes to match spectra with candidate cross-linked peptides have been described (Schilling et al., 2003; de Koning et al., 2006; Gao et al., 2006; Maiolica et al., 2007; Heymann et al., 2008; Nadeau et al., 2008; Rinner et al., 2008; Singh et al., 2008; Chu et al., 2010; McIlwain et al., 2010; Petrotchenko and Borchers, 2010; Xu et al., 2010) and recently been reviewed (Leitner et al., 2010; Singh et al., 2010). Nevertheless, this is an area of ongoing developments not least because of a second challenge: determining the confidence of a match. False identifications of linear peptides have been reduced through manual interrogation of peptide-spectrum matches, by applying filters created using a training data set (Eng et al., 1994), using probabilistic approaches (Perkins et al., 1999; Nesvizhskii et al., 2003; Sadygov and Yates, 2003), relying on machine learning (Käll et al., 2007), and using the target-decoy approach, combining the ordinary (target) database usually with an inverted (decoy) database (Moore et al., 2002; Kislinger et al., 2003; Peng et al., 2003; Sennels et al., 2009). Following these experiences with linear peptides, the false positive rate of database searches for cross-linked peptides has been estimated by using the target-decoy method (Maiolica et al., 2007; Rinner et al., 2008), relying on a decoy database or using a false mass for the cross-linker, and manual interrogation following a decision tree (Chen et al., 2010).

3. RNA Polymerase II complexes

3.1. RNA Polymerase II core complex – benchmarking cross-linking/mass spectrometry

The analyses of protein complexes such as our success in using cross-linking/mass spectrometry to guide the engineering of a crystallisable *H. sapiens* Ndc80$^{\text{bonsai}}$ complex (Maiolica et al., 2007; Ciferri et al., 2008) demonstrated that the technology is principally of value. However, a detailed analysis of the data's accuracy was not possible, as all of these studies reported a relatively small number of linkage sites (10-25 at best). We therefore targeted recently (Chen et al., 2010) a large multi-protein complex, the *S. cerevisiae* RNA polymerase II (Pol II), for which a crystal structure had been deposited (PDB 1WCM) (Armache et al., 2005) and could be used as a reference to check the quality of cross-link data.

Purified Pol II complexes (12 subunits, 513 kDa) cross-linked readily, as could be seen from the change in protein bands under denaturating gel electrophoresis before and after cross-linking (Figure 2A) (Chen et al., 2010). We used the cross-linker Bis (sulpho-succinimidyl) suberate (BS3, Thermo Fisher), which reacts with primary amines in lysine side chains and protein N-termini. Cross-linking did not lead to extensive aggregation of complexes, as could be seen from native gel electrophoresis (Figure 2B). 30 μg of Pol II were subjected to our analysis: cross-linking, gel electrophoresis to isolate monomeric complexes, trypsin digestion, fractionation of peptides by strong-cation exchange chromatography, liquid chromatography-mass spectrometry in a high-high strategy, and finally database searching to identify the cross-linked peptides. In summary, 429 fragmentation spectra matched to cross-linked peptides covering 146 unique linkage pairs. From this data, 106 linkage pairs were obtained for which distance data could be extracted from the crystal structure of Pol II. Following a decision tree the data supporting the 106 linkages was classified for its quality, leading to 80 higher-confidence and 26 lower-confidence cross-links.

The distance distribution for alpha-carbon pairs of cross-linked lysines was clearly different from a random set of lysine pairs selected from the crystal structure (Figure 2C). Based on this comparison, two arguments could be made for the accuracy of the cross-link/mass spectrometry data. First, the observed distance distribution was very unlikely to be a random result (P-value of 3×10^{-87}). Second, the observed distribution looked plausible when considering the length of a lysine side chain to be 6-6.5 Angstrom, the length of the cross-linker in full extension to be 11.4 Angstrom, and an experimental error for this crystal structure of 1-1.5 Angstrom for surface residues (as estimated from the crystallographic B-factor). Adding all these together would predict the majority of cross-links to report lysine pairs whose alpha-carbons are closer than 27.4 Angstrom in the crystal structure. This was indeed the case for 93% of the data.

With six of seven cross-links above 27 Angstrom, the cross-links of longest length tended to fall into the mobile clamp domain of Pol II. These long distance cross-links could therefore be rationalised as capturing conformations of Pol II possible in solution. A single cross-link supposedly bridged residues whose alpha-carbons were positioned nearly 60 Angstrom apart in the crystal structure. Dense packing of protein separates the residues according to the crystal structure, which makes conformational changes unlikely to bring these two residues into close enough proximity for cross-linking. Furthermore, the cross-link distance fell into the broad maximum of the randomly selected pairs and the data supporting the cross-link had been classified as being of lower confidence. We hence concluded that this single cross-link among the 106 observed cross-links was a false positive, suggesting a false positive rate of less than 1% when combining higher and lower confidence data.

3.2. RNA Polymerase II-TFIIF – expanding a stable complex core towards its more elusive periphery

We next analysed the complex of Pol II with transcription factor IIF (TFIIF), comprising 15 subunits with a total molecular weight of 670 kDa, including the three subunits of TFIIF: Tfg1, Tfg2, and Tfg3 (Chen et al., 2010). A crystal structure of the Pol II complex and crystal structures for three domains of human TFIIF subunits composed the structural pre-knowledge. Using 200 μg of purified complex and following the same strategy as outlined above we identified 402 linkage sites within TFIIF or between Pol II and TFIIF. Cross-links within Pol II were observed but not evaluated. Using a decision tree as above, 224 higher-confidence cross-links were selected and used for model building. The data was summarised in form of a linkage map (Figure 3A). This linkage map of the Pol II-TFIIF complex supported the validity of homology models for three TFIIF domains, provided a reciprocal footprint of TFIIF on Pol II and of Pol II on TFIIF at peptide resolution, and led to the docking of a homology model for the Tfg1/Tfg2 dimerisation domain of TFIIF with the Pol II crystal structure.

Homology modelling can be used to infer the structure of a novel protein or domain if the structure of a related protein or domain has already been determined at high resolution. Structures for the human winged-helix domains of Tfg1 and Tfg2 (Groft et al., 1998; Kamada et al., 2001) as well as for the Tfg1/Tfg2 dimerisation domain (Gaiser et al., 2000) had been solved and could be used as templates for homology modelling. The sequence alignment for the dimerization domain from *H. sapiens* and *S. cerevisiae* was not unambiguous, however, leaving an element of uncertainty on aspects of the model. Cross-links cannot currently be used to assist homology modelling. However the homology models obtained can be challenged by the experimental data. Indeed, the cross-link data for TFIIF was incorporated into the finished homology models in order to test if the model satisfied the experimental constraints.

The cross-link data between TFIIF and Pol II revealed the interaction regions between TFIIF and Pol II and located TFIIF on the Pol II surface (Figure 3B). Cross-link sites in Pol II were colour coded by TFIIF region to visualize the footprint of individual TFIIF regions on the surface of Pol II. The data revealed distinct areas on Pol II that interact with the three TFIIF subunits. The position of different Tfg1 and Tfg2 regions could be followed in detail. For Tfg1, the N-terminal tail, dimerization domain and charged region were positioned on Pol II. For Tfg2, the dimerization domain, linker, and winged-helix domain were positioned on Pol II. Tfg3 is located in a region on Pol II, that in crystal structures of Pol II is occupied by other transcription factors. This may indicate the importance of studying as fully assembled complexes as possible, because Pol II-TFIIF is only a sub-complex of the pre-initiation complex. Similarly, the C-terminal region of Tfg2 including the winged-helix domain displayed a number of alternative binding positions, some of which are not possible in the pre-initiation complex.

The mutually exclusive binding positions for the Tfg2 C-terminal region as revealed by cross-linking/MS demonstrated an ability of this technology to captured dynamic situations in protein complexes that is at the same time exciting and challenging. The fact that dynamic situations can be revealed by cross-linking/MS is exciting. The challenge is in the fact that the data of all different states of a complex or protein are superimposed. Utilizing the Pol II crystal structure has permitted the disentanglement of the overlying cross-link data and this has revealed the dynamic aspects of TFIIF binding.

Interestingly, the Tfg1 winged-helix domain was not found to link to any region of the Pol II-TFIIF complex other than the domain itself. As this domain is being held close to the rest of the complex by a linker region it will have as a minimum random contacts with surfaces of the rest of the complex. These not being sufficient to lead to observable cross-links may indicate that transient interactions need to be long enough to provide the time required for the cross-linking reaction to take place and that the interactions have to represent a significant fraction of the population. In other words, a lower threshold for the stability of structures exists for them to be captured by cross-linking. Indeed, a recent study found a protein complexes with K(D) ~ 25 μM to cross-linked specifically while another complex with K(D) 100-300 μM did not, indicating the limit for cross-linking to be somewhere in this affinity range (Madler et al., 2010b).

The interaction between the Tfg1/Tfg2 dimerisation domain and Pol II surpassed this threshold and was observed by formation of numerous cross-links. This allowed docking of the domain and Pol II (Figure 3C). As for the Tfg2 winged-helix domain, positional ambiguity resulted from cross-link data that could not be satisfied by a single binding mode. Taken together, this analysis found TFIIF binding to Pol II in multiple modes that possibly exchange in a dynamic fashion. The lessons learned from these analyses of two Pol II complexes are presented below, with regards to integrating of cross-linking/MS data into the modelling process as well as planning structural studies that utilize this technology.

4. Challenges of modelling when using cross-link data

The concept of a distance constraint is not new to modelling. Distance constraints are provided in large quantities for small proteins or domains in NMR (Nilges et al., 1997). Also in small quantities for larger proteins and multi-protein complexes by other biophysical techniques, typically after introducing specific probes, e.g. spin labels in EPR (for review of low resolution methods and modelling (Venselaar et al., 2010)). Cross-link derived constraints are different from NMR data in being sparse and long distance. Even at "zero length" the length of the cross-linked side chains add to over 10 Angstrom between the alpha-carbons of the linked residues. However, cross-linking yields constraints more plentiful and easier than any low-resolution biophysical method. A proper treatment of low-resolution distance constraints is now indicated for modelling. The following points at least should be considered when integrating cross-link data into modelling software.

4.1. Experimental data can be ambiguous

As an experimental method, cross-linking/MS will yield data with an experimental error. In our benchmarking experiment using the Pol II we found an experimental error of less than 1% with respect to miss-assigned linkages when combining high and low confidence data. None of the high confidence data proved to be incorrect, indicating that cross-linking/MS can yield unambiguous data. Nevertheless, there may also be value in lower confidence constrains, to reflect underrepresented conformers/structures or provide additional constraints for modelling. Thus, it would be desirable if modelling software could use constraint information while simultaneously taking their confidence level into account. In addition to this experimental error there is also positional ambiguity, when the site of linkage cannot be narrowed to a single residue but only a stretch of residues due to lack of fragmentation information. Alternatively, the same peptide sequence might be found in more than one location of a protein sequence or in more than one protein of a complex. This is particularly likely when detecting short peptides as partners in cross-linked peptides. Any modelling software should be able to deal with this ambiguity and reward models that satisfy at least one of the constraint alternatives.

4.2. From distance constraint to distance restraint

In first approximation, the distance constraint for the position of alpha-carbons in two cross-linked residues can be calculated by adding the length of the spacer in the cross-linker and the length of the linked side chains. This neglects, however, the dynamic behaviour of molecules in solution. Due to bond rotations and vibrations in the spacer, the cross-linker will sample a certain length distribution with the fully extended conformation being only one of many possible states. This has been modelled for a number of cross-linkers and a shorter "effective" length been proposed (Green et al., 2001). However, the protein(s) will also sample their conformational space in solution. The extent of residue movements will be protein and position dependent and as such is currently unpredictable. The influence of protein vibrations or conformational flexibility is likely to be of significantly larger importance than that of the cross-linker, especially when analysing large proteins or multi-protein complexes. Experimental data such as obtained for the Pol II may offer a heuristic solution to this problem. The amino acid pairs that were cross-linked did not spread equally over the range defined by the distance constraint. This suggests the possibility of using distance restraints instead of constraints. Using data obtained with cross-linkers of different length will improve the distance restraint by providing information on lower limits.

4.3. Cross-linking is undemocratic

Cross-linking/MS will result in more data for some parts of a protein or complex than in others. Several reasons account for this and a number of approaches can be taken to obtain at least partial remedy of the undemocratic nature of the approach. Firstly, cross-linking requires reactive sites in the protein(s) to be available, accessible, and in linkable geometry. Lysines have been consequently the target of choice, as they tend to be plentiful, accessible on the surface of proteins, and react with high specificity with N-hydroxysuccinimide cross-linkers (note that side reactions with serines, threonines and tyrosines have been observed

(Leavell et al., 2004; Swaim et al., 2004; Kalkhof and Sinz, 2008; Madler et al., 2009; Madler et al., 2010a)). The distribution of lysine residues on the surface of proteins is, however, not even. Consequently, constraint data will vary in its coverage of a structure and be particularly scarce in hydrophobic regions such hydrophobic cores or transmembrane regions. Experimentally, this can be addressed by targeting different residue pairs from lysine-lysine such as lysine-aspartate/glutamate, lysine-cysteine or cysteine-cysteine, all for which commercial cross-linking reagents are available. Also, use of photoactivatable linkers and even photoactivatable amino acid analogues such as azido-methionine or azido-leucine (Suchanek et al., 2005) or arginine-arginine cross-linkers (Zhang et al., 2008) are being explored. However, the patchy nature of cross-link data means that modelling will usually require additional structural data.

4.4. Absence of data is inconclusive but possibly suggestive

Not all cross-linked residues will actually be detected in cross-linked peptides. Contributing factors are the size limitation of standard mass spectrometers, masking of peaks by background and loss of hydrophobic peptides during the sample preparation. The apparent absence of an individual cross-link between two residues cannot justify the assumption that the two residues are not proximal. However, when considering groups of cross-links absence of data might still be informative. For example, the presence of multiple cross-links between two regions A and B and the presence of multiple cross-links between two other regions C and D would indicate that cross-links in all four regions were principally observable. Absence of cross-links between A/B and C/D would then indicate a lower probability of A/B and C/D to be proximal.

4.5. Population data

The protein(s) under investigation may be heterogeneous in a number of different ways. They will sample the conformational space available to them under the experimental conditions, i.e. not exist in one static structure. They may differ furthermore in their modification states, e.g. the presence and absence of a particular phosphorylation that may induce a large conformational change in a protein. On- and off-rates of subunits may lead to the presence of multiple forms of a complex. A particular strength of cross-linking is its ability to work with crude starting material. This means that even multiple forms of a complex may be present such as assembly intermediates, fully assembled complexes but with different subunit composition to conduct specialised functions, or complexes at different processing states, differing in transient factors or conformations. Cross-linking will reflect this diversity to some extend and not just reflect a single static structure. No single model may therefore fulfil all constraints derived from cross-link data. Instead, modelling needs to create an ensemble of structures that relate to each other by conformational or compositional changes. Ultimately, this means that modelling will need to move closer to the structural reality by including dynamic aspects of proteins and protein complexes.

4.6. Artefacts

Cross-linking modifies proteins and could principally result in structural artefacts in a number of ways, none of which has been shown to occur experimentally so far. An investigated protein could be caught in a very rare conformation, with further cross-linking events possibly exasperating the effect to create and therefore also reflect an otherwise impossible conformation. Similarly, a randomly passing protein might be caught to create a fusion that is not reflecting a functional protein-protein interaction. As cross-linking/MS provides data on populations these rare events have so far remained hidden among more frequent and less artefactual products of cross-linking. Cross-linking might furthermore trigger a conformation change that either is physiological but would normally require a stimulus such as binding of a co-factor or that is artefactual. At the current, possibly still initial state of comparing cross-linking/MS data with high-resolution structures an extensive agreement between the two methods has been testified (Maiolica et al., 2007; Rinner et al., 2008; Chen et al., 2010). This argues at least against induced conformational changes being a frequent artefact of cross-linking. Finally, aggregates of proteins that arise from the high concentrations of protein solutions often used for cross-linking may be cross-linked and lead to artefactual protein-protein contacts. This can and has to be controlled for in each analysis by checking the analysed sample for cross-linked aggregates, for example using native gel electrophoresis (Chen et al., 2010). Besides these cross-linking artefacts there is also the possibility of erroneous data interpretation. This is possible for example during the data base search (Section 2.4) or when considering ambiguous data during modelling (Section 4.1) and needs to be carefully controlled for. However, all currently available data indicate that cross-linking/MS is a reliable source of structural information.

4.7. Quantitative information

Cross-linking/mass spectrometry can provide quantitative data on the structure of proteins and complexes. It is tempting to speculate that this may pave the way for modelling to simulate dynamic structures. Quantitative information can in principle be obtained by comparing mass spectrometric signal intensities of different cross-links within an experiment and for identical cross-links in different experiments.

Comparing the yield for links across different experiments can be done by quantitative proteomics, relying on stable isotope labelling or label free approaches. Signal intensities for a cross-linked peptide or set of cross-linked peptides can be compared across different analyses. Alternatively, stable isotopes can be used to encode in a single analysis (typically leading to higher accuracy) the origin of peptides to different experiments to allow using their relative peak intensities for quantitation. This approach is well established in proteomics and relies on incorporating stable isotopes for example by amino acids in cell culture (SILAC) (Ong et al., 2002; Blagoev et al., 2003; Ong et al., 2003). Isotope labelled chemical modifiers that react, for example, with the N-terminus of all peptides (iTRAQ and TMT) can be used alternatively. In addition, a pair of light and heavy labelled cross-linker might be used (Petrotchenko et al., 2005; Chu et al., 2006a; Ihling et al., 2006). For example, a protein could be cross-linked with a light cross-linker under one condition and with the heavy version of the same cross-linker under a different condition that may, for example, change the protein conformation. Mixing the two proteins and then analysing the cross-linked peptides by mass spectrometry will create signal pairs for every cross-linked peptide. The peak ratio of the pair gives the relative yield of this cross-link under the two conditions and thus reveals to what extend the cross-linked sites are affected by the condition change.

Comparing the yield of different cross-links within an experiment is more challenging. Firstly, yield differences result from interplay of several environmental factors such as accessibility and reactivity of both sites, their relative position to each other in terms of orientation of side chains and possible obstructions, and conformational flexibility bringing sites into sufficient proximity possibly with only limited occurrence. In addition, peptides differ in their respective “flyability”, the intensity by which a given peptide is observed in a mass spectrometer. Different peptides can therefore not be compared individually with one another. This means that cross-linking does not provide any information on the abundance of an observed proximity. However, the “flyability” differences between peptides can be statistically averaged when comparing groups of peptides. This strategy is used when comparing the concentration of different proteins in the same sample (Rappsilber et al., 2002; Ishihama et al., 2005). Comparing, for example, groups of conformation-specific cross-links may well indicate the relative ratio by which the conformations occur.

5. Conclusion

After more than a decade of method developments and proof-of-concept studies, cross-linking/mass spectrometry is nearing the end of adolescence. The analysis of several multi-protein complexes shows this technology to yield rich constraint data that expedites the modelling process of proteins and multi-protein complex. The technology itself may still need further consolidation, namely the development of a user-friendly and powerful database search engine. Also, more studies are needed to develop cross-linking/mass spectrometry into a generally accepted method for deriving structural information on protein complexes. However, modellers should prepare now for the arrival of large amounts of experimental data and fresh impetus to integrated structural biology from cross-linking/MS. The ease by which experimental data can be generated will ultimately have to be matched by the ease with which modelling tools can be operated, to allow integrated structural biology to spread throughout diverse fields in life sciences.

Acknowledgments

I thank the members of the Rappsilber lab and Michael Nilges for discussion and Hanno Steen for critically reading the manuscript. This work was supported by a European Community Marie Curie Excellence Grant and the Wellcome Trust. JR is a Wellcome Trust Senior Research Fellow.

References

Aebersold R, Mann M. Mass spectrometry-based proteomics. Nature. 2003; 422:198–207. [PubMed: 12634793]

Alber F, Dokudovskaya S, Veenhoff LM, Zhang W, Kipper J, Devos D, Suprapto A, Karni-Schmidt O, Williams R, Chait BT, Rout MP, Sali A. Determining the architectures of macromolecular assemblies. Nature. 2007a; 450:683–94. [PubMed: 18046405]

Alber F, Dokudovskaya S, Veenhoff LM, Zhang W, Kipper J, Devos D, Suprapto A, Karni-Schmidt O, Williams R, Chait BT, Sali A, Rout MP. The molecular architecture of the nuclear pore complex. Nature. 2007b; 450:695–701. [PubMed: 18046406]

Armache KJ, Mitterweger S, Meinhart A, Cramer P. Structures of complete RNA polymerase II and its subcomplex, Rpb4/7. J Biol Chem. 2005; 280:7131–4. [PubMed: 15591044]

Back JW, de Jong L, Muijsers AO, de Koster CG. Chemical cross-linking and mass spectrometry for protein structural modeling. J Mol Biol. 2003; 331:303–13. [PubMed: 12888339]

Back JW, Hartog AF, Dekker HL, Muijsers AO, de Koning LJ, de Jong L. A new crosslinker for mass spectrometric analysis of the quaternary structure of protein complexes. J Am Soc Mass Spectrom. 2001; 12:222–7. [PubMed: 11212007]

Back JW, Notenboom V, de Koning LJ, Muijsers AO, Sixma TK, de Koster CG, de Jong L. Identification of cross-linked peptides for protein interaction studies using mass spectrometry and 18O labeling. Anal Chem. 2002; 74:4417–22. [PubMed: 12236350]

Benesch JL, Robinson CV. Mass spectrometry of macromolecular assemblies: preservation and dissociation. Curr Opin Struct Biol. 2006; 16:245–51. [PubMed: 16563743]

Benesch JL, Ruotolo BT, Simmons DA, Robinson CV. Protein complexes in the gas phase: technology for structural genomics and proteomics. Chem Rev. 2007; 107:3544–67. [PubMed: 17649985]

Bich C, Maedler S, Chiesa K, DeGiacomo F, Bogliotti N, Zenobi R. Reactivity and applications of new amine reactive cross-linkers for mass spectrometric detection of protein-protein complexes. Anal Chem. 2010; 82:172–9. [PubMed: 19994840]

Biemann K. Contributions of mass spectrometry to peptide and protein structure. Biomed Environ Mass Spectrom. 1988; 16:99–111. [PubMed: 3072035]

Blagoev B, Kratchmarova I, Ong SE, Nielsen M, Foster LJ, Mann M. A proteomics strategy to elucidate functional protein-protein interactions applied to EGF signaling. Nat Biotechnol. 2003; 21:315–8. [PubMed: 12577067]

Chen ZA, Jawhari A, Fischer L, Buchen C, Tahir S, Kamenski T, Rasmussen M, Lariviere L, Bukowski-Wills JC, Nilges M, Cramer P, Rappsilber J. Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry. Embo J. 2010; 29:717–26. [PubMed: 20094031]

Chowdhury SM, Munske GR, Tang X, Bruce JE. Collisionally activated dissociation and electron capture dissociation of several mass spectrometry-identifiable chemical cross-linkers. Anal Chem. 2006; 78:8183–93. [PubMed: 17165806]

Chowdhury SM, Du X, Tolic N, Wu S, Moore RJ, Mayer MU, Smith RD, Adkins JN. Identification of cross-linked peptides after click-based enrichment using sequential collision-induced dissociation and electron transfer dissociation tandem mass spectrometry. Anal Chem. 2009; 81:5524–32. [PubMed: 19496583]

Chu F, Mahrus S, Craik CS, Burlingame AL. Isotope-coded and affinity-tagged cross-linking (ICATXL): an efficient strategy to probe protein interaction surfaces. J Am Chem Soc. 2006a; 128:10362–3. [PubMed: 16895390]

Chu F, Baker PR, Burlingame AL, Chalkley RJ. Finding chimeras: a bioinformatics strategy for identification of cross-linked peptides. Mol Cell Proteomics. 2010; 9:25–31. [PubMed: 19809093]

Chu F, Maynard JC, Chiosis G, Nicchitta CV, Burlingame AL. Identification of novel quaternary domain interactions in the Hsp90 chaperone, GRP94. Protein Sci. 2006b; 15:1260–9. [PubMed: 16731965]

Chu F, Shan SO, Moustakas DT, Alber F, Egea PF, Stroud RM, Walter P, Burlingame AL. Unraveling the interface of signal recognition particle and its receptor by using chemical cross-linking and tandem mass spectrometry. Proc Natl Acad Sci U S A. 2004; 101:16454–9. [PubMed: 15546976]

Ciferri C, Pasqualato S, Screpanti E, Varetti G, Santaguida S, Dos Reis G, Maiolica A, Polka J, De Luca JG, De Wulf P, Salek M, Rappsilber J, Moores CA, Salmon ED, Musacchio A. Implications for kinetochore-microtubule attachment from the structure of an engineered Ndc80 complex. Cell. 2008; 133:427–39. [PubMed: 18455984]

Clegg C, Hayes D. Identification of neighbouring proteins in the ribosomes of Escherichia coli. A topographical study with the cross-linking reagent dimethyl suberimidate. Eur J Biochem. 1974; 42:21–8. [PubMed: 4598101]

de Koning LJ, Kasper PT, Back JW, Nessen MA, Vanrobaeys F, Van Beeumen J, Gherardi E, de Koster CG, de Jong L. Computer-assisted mass spectrometric analysis of naturally occurring and artificially introduced cross-links in proteins and protein complexes. FEBS J. 2006; 273:281–91. [PubMed: 16403016]

Denison C, Kodadek T. Toward a general chemical method for rapidly mapping multi-protein complexes. J Proteome Res. 2004; 3:417–25. [PubMed: 15253422]

Dimova K, Kalkhof S, Pottratz I, Ihling C, Rodriguez-Castaneda F, Liepold T, Griesinger C, Brose N, Sinz A, Jahn O. Structural insights into the calmodulin-Munc13 interaction obtained by cross-linking and mass spectrometry. Biochemistry. 2009; 48:5908–21. [PubMed: 19492809]

Dreiocker F, Muller MQ, Sinz A, Schafer M. Collision-induced dissociative chemical cross-linking reagent for protein structure characterization: applied Edman chemistry in the gas phase. J Mass Spectrom. 2010; 45:178–89. [PubMed: 19950134]

Eng JK, McCormack AL, Yates JR 3rd. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J. Am. Soc. Mass Spectrom. 1994; 5:976–989.

Fu CY, Uetrecht C, Kang S, Morais MC, Heck AJ, Walter MR, Prevelige PE Jr. A docking model based on mass spectrometric and biochemical data describes phage packaging motor incorporation. Mol Cell Proteomics. 2010

Gaiser F, Tan S, Richmond TJ. Novel dimerization fold of RAP30/RAP74 in human TFIIF at 1.7 A resolution. J Mol Biol. 2000; 302:1119–27. [PubMed: 11183778]

Gao Q, Xue S, Doneanu CE, Shaffer SA, Goodlett DR, Nelson SD. Pro-CrossLink. Software tool for protein cross-linking and mass spectrometry. Anal Chem. 2006; 78:2145–9. [PubMed: 16579592]

Gardner MW, Vasicek LA, Shabbir S, Anslyn EV, Brodbelt JS. Chromogenic cross-linker for the characterization of protein structure by infrared multiphoton dissociation mass spectrometry. Anal Chem. 2008; 80:4807–19. [PubMed: 18517224]

Gaucher SP, Hadi MZ, Young MM. Influence of crosslinker identity and position on gas-phase dissociation of Lys-Lys crosslinked peptides. J Am Soc Mass Spectrom. 2006; 17:395–405. [PubMed: 16443364]

Gomes AF, Gozzo FC. Chemical cross-linking with a diazirine photoactivatable cross-linker investigated by MALDI- and ESI-MS/MS. J Mass Spectrom. 2010; 45:892–9. [PubMed: 20635431]

Green NS, Reisler E, Houk KN. Quantitative evaluation of the lengths of homobifunctional protein cross-linking reagents used as molecular rulers. Protein Sci. 2001; 10:1293–304. [PubMed: 11420431]

Groft CM, Uljon SN, Wang R, Werner MH. Structural homology between the Rap30 DNA-binding domain and linker histone H5: implications for preinitiation complex assembly. Proc Natl Acad Sci U S A. 1998; 95:9117–22. [PubMed: 9689043]

Heck AJ. Native mass spectrometry: a bridge between interactomics and structural biology. Nat Methods. 2008; 5:927–33. [PubMed: 18974734]

Heymann M, Paramelle D, Subra G, Forest E, Martinez J, Geourjon C, Deleage G. MSX-3D: a tool to validate 3D protein models using mass spectrometry. Bioinformatics. 2008; 24:2782–3. [PubMed: 18826958]

Iglesias AH, Santos LF, Gozzo FC. Collision-induced dissociation of Lys-Lys intramolecular crosslinked peptides. J Am Soc Mass Spectrom. 2009; 20:557–66. [PubMed: 19138533]

Iglesias AH, Santos LF, Gozzo FC. Identification of cross-linked peptides by high-resolution precursor ion scan. Anal Chem. 2010; 82:909–16. [PubMed: 20063856]

Ihling C, Schmidt A, Kalkhof S, Schulz DM, Stingl C, Mechtler K, Haack M, Beck-Sickinger AG, Cooper DM, Sinz A. Isotope-labeled cross-linkers and Fourier transform ion cyclotron resonance mass spectrometry for structural analysis of a protein/peptide complex. J Am Soc Mass Spectrom. 2006; 17:1100–13. [PubMed: 16750914]

Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, Mann M. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. Mol Cell Proteomics. 2005; 4:1265–72. [PubMed: 15958392]

Jacobsen RB, Sale KL, Ayson MJ, Novak P, Hong J, Lane P, Wood NL, Kruppa GH, Young MM, Schoeniger JS. Structure and dynamics of dark-state bovine rhodopsin revealed by chemical cross-linking and high-resolution mass spectrometry. Protein Sci. 2006; 15:1303–17. [PubMed: 16731966]

Kalkhof S, Sinz A. Chances and pitfalls of chemical cross-linking with amine-reactive N-hydroxysuccinimide esters. Anal Bioanal Chem. 2008; 392:305–12. [PubMed: 18724398]

Käll L, Canterbury JD, Weston J, Noble WS, MacCoss MJ. Semi-supervised learning for peptide identification from shotgun proteomics datasets. Nat Methods. 2007; 4:923–5. [PubMed: 17952086]

Kamada K, De Angelis J, Roeder RG, Burley SK. Crystal structure of the C-terminal domain of the RAP74 subunit of human transcription factor IIF. Proc Natl Acad Sci U S A. 2001; 98:3115–20. [PubMed: 11248041]

Kang S, Mou L, Lanman J, Velu S, Brouillette WJ, Prevelige PE Jr. Synthesis of biotin-tagged chemical cross-linkers and their applications for mass spectrometry. Rapid Commun Mass Spectrom. 2009; 23:1719–26. [PubMed: 19412923]

Karaca E, Melquiond AS, de Vries SJ, Kastritis PL, Bonvin AM. Building macromolecular assemblies by information-driven docking: introducing the HADDOCK multi-body docking server. Mol Cell Proteomics. 2010

Kasper PT, Back JW, Vitale M, Hartog AF, Roseboom W, de Koning LJ, van Maarseveen JH, Muijsers AO, de Koster CG, de Jong L. An aptly positioned azido group in the spacer of a protein cross-linker for facile mapping of lysines in close proximity. Chembiochem. 2007; 8:1281–92. [PubMed: 17600791]

Kislinger T, Rahman K, Radulovic D, Cox B, Rossant J, Emili A. PRISM, a generic large scale proteomic investigation strategy for mammals. Mol Cell Proteomics. 2003; 2:96–106. [PubMed: 12644571]

Krauth F, Ihling CH, Ruttinger HH, Sinz A. Heterobifunctional isotope-labeled amine-reactive photo-cross-linker for structural investigation of proteins by matrix-assisted laser desorption/ionization tandem time-of-flight and electrospray ionization LTQ-Orbitrap mass spectrometry. Rapid Commun Mass Spectrom. 2009; 23:2811–8. [PubMed: 19653199]

Leavell MD, Novak P, Behrens CR, Schoeniger JS, Kruppa GH. Strategy for selective chemical cross-linking of tyrosine and lysine residues. J Am Soc Mass Spectrom. 2004; 15:1604–11. [PubMed: 15519227]

Leitner A, Walzthoeni T, Kahraman A, Herzog F, Rinner O, Beck M, Aebersold R. Probing native protein structures by chemical cross-linking, mass spectrometry and bioinformatics. Mol Cell Proteomics. 2010

Liu F, Goshe MB. Combinatorial Electrostatic Collision-Induced Dissociative Chemical Cross-linking Reagents for Probing Protein Surface Topology. Anal Chem. 2010

Lu Y, Tanasova M, Borhan B, Reid GE. Ionic reagent for controlling the gas-phase fragmentation reactions of cross-linked peptides. Anal Chem. 2008; 80:9279–87. [PubMed: 19551991]

Madler S, Gschwind S, Zenobi R. Role of arginine in chemical cross-linking with N-hydroxysuccinimide esters. Anal Biochem. 2010a; 398:123–5. [PubMed: 19931213]

Madler S, Bich C, Touboul D, Zenobi R. Chemical cross-linking with NHS esters: a systematic study on amino acid reactivities. J Mass Spectrom. 2009; 44:694–706. [PubMed: 19132714]

Madler S, Seitz M, Robinson J, Zenobi R. Does chemical cross-linking with NHS esters reflect the chemical equilibrium of protein-protein noncovalent interactions in solution? J Am Soc Mass Spectrom. 2010b; 21:1775–83. [PubMed: 20708949]

Maiolica A, Cittaro D, Borsotti D, Sennels L, Ciferri C, Tarricone C, Musacchio A, Rappsilber J. Structural analysis of multiprotein complexes by cross-linking, mass spectrometry, and database searching. Mol Cell Proteomics. 2007; 6:2200–11. [PubMed: 17921176]

McIlwain S, Draghicescu P, Singh P, Goodlett DR, Noble WS. Detecting cross-linked peptides by searching against a database of cross-linked peptide pairs. J Proteome Res. 2010; 9:2488–95. [PubMed: 20349954]

Moore RE, Young MK, Lee TD. Qscore: an algorithm for evaluating SEQUEST database search results. J Am Soc Mass Spectrom. 2002; 13:378–86. [PubMed: 11951976]

Mouradov D, King G, Ross IL, Forwood JK, Hume DA, Sinz A, Martin JL, Kobe B, Huber T. Protein structure determination using a combination of cross-linking, mass spectrometry, and molecular modeling. Methods Mol Biol. 2008; 426:459–74. [PubMed: 18542884]

Muller MQ, de Koning LJ, Schmidt A, Ihling C, Syha Y, Rau O, Mechtler K, Schubert-Zsilavecz M, Sinz A. An innovative method to study target protein-drug interactions by mass spectrometry. J Med Chem. 2009; 52:2875–9. [PubMed: 19379014]

Nadeau OW, Wyckoff GJ, Paschall JE, Artigues A, Sage J, Villar MT, Carlson GM. CrossSearch, a user-friendly search engine for detecting chemically cross-linked peptides in conjugated proteins. Mol Cell Proteomics. 2008; 7:739–49. [PubMed: 18281724]

Nessen MA, Kramer G, Back J, Baskin JM, Smeenk LE, de Koning LJ, van Maarseveen JH, de Jong L, Bertozzi CR, Hiemstra H, de Koster CG. Selective enrichment of azide-containing peptides from complex mixtures. J Proteome Res. 2009; 8:3702–11. [PubMed: 19402736]

Nesvizhskii AI, Keller A, Kolker E, Aebersold R. A statistical model for identifying proteins by tandem mass spectrometry. Anal Chem. 2003; 75:4646–58. [PubMed: 14632076]

Nilges M, Macias MJ, O'Donoghue SI, Oschkinat H. Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from beta-spectrin. J Mol Biol. 1997; 269:408–22. [PubMed: 9199409]

Ong SE, Foster LJ, Mann M. Mass spectrometric-based approaches in quantitative proteomics. Methods. 2003; 29:124–30. [PubMed: 12606218]

Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics. 2002; 1:376–86. [PubMed: 12118079]

Paramelle D, Cantel S, Enjalbal C, Amblard M, Forest E, Heymann M, Geourjon C, Martinez J, Subra G. A new generation of cross-linkers for selective detection by MALDI MS. Proteomics. 2009; 9:5384–8. [PubMed: 19902427]

Peng J, Elias JE, Thoreen CC, Licklider LJ, Gygi SP. Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. J Proteome Res. 2003; 2:43–50. [PubMed: 12643542]

Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis. 1999; 20:3551–67. [PubMed: 10612281]

Petrotchenko EV, Borchers CH. ICC-CLASS: isotopically-coded cleavable crosslinking analysis software suite. BMC Bioinformatics. 2010; 11:64. [PubMed: 20109223]

Petrotchenko EV, Olkhovik VK, Borchers CH. Isotopically coded cleavable cross-linker for studying protein-protein interaction and protein complexes. Mol Cell Proteomics. 2005; 4:1167–79. [PubMed: 15901824]

Petrotchenko EV, Serpa JJ, Borchers CH. Use of a combination of isotopically coded cross-linkers and isotopically coded N-terminal modification reagents for selective identification of inter-peptide crosslinks. Anal Chem. 2010; 82:817–23. [PubMed: 20050626]

Petrotchenko EV, Xiao K, Cable J, Chen Y, Dokholyan NV, Borchers CH. BiPS, a photocleavable, isotopically coded, fluorescent cross-linker for structural proteomics. Mol Cell Proteomics. 2009; 8:273–86. [PubMed: 18838738]

Pimenova T, Nazabal A, Roschitzki B, Seebacher J, Rinner O, Zenobi R. Epitope mapping on bovine prion protein using chemical cross-linking and mass spectrometry. J Mass Spectrom. 2008; 43:185–95. [PubMed: 17924399]

Rappsilber J, Siniossoglou S, Hurt EC, Mann M. A generic strategy to analyze the spatial organization of multi-protein complexes by cross-linking and mass spectrometry. Anal Chem. 2000; 72:267–75. [PubMed: 10658319]

Rappsilber J, Ryder U, Lamond AI, Mann M. Large-scale proteomic analysis of the human spliceosome. Genome Res. 2002; 12:1231–45. [PubMed: 12176931]

Rinner O, Seebacher J, Walzthoeni T, Mueller LN, Beck M, Schmidt A, Mueller M, Aebersold R. Identification of cross-linked peptides from large sequence databases. Nat Methods. 2008; 5:315–8. [PubMed: 18327264]

Roepstorff P, Fohlman J. Proposed nomenclature for sequence ions. Biomed Mass Spectrom. 1984; 11:601. [PubMed: 6525415]

Sadygov RG, Yates JR 3rd. A hypergeometric probability model for protein identification and validation using tandem mass spectral data and protein sequence databases. Anal Chem. 2003; 75:3792–8. [PubMed: 14572045]

Schilling B, Row RH, Gibson BW, Guo X, Young MM. MS2Assign, automated assignment and nomenclature of tandem mass spectra of chemically crosslinked peptides. J Am Soc Mass Spectrom. 2003; 14:834–50. [PubMed: 12892908]

Schulz DM, Kalkhof S, Schmidt A, Ihling C, Stingl C, Mechtler K, Zschornig O, Sinz A. Annexin A2/P11 interaction: new insights into annexin A2 tetramer structure by chemical crosslinking, high-resolution mass spectrometry, and computational modeling. Proteins. 2007; 69:254–69. [PubMed: 17607745]

Seebacher J, Mallick P, Zhang N, Eddes JS, Aebersold R, Gelb MH. Protein cross-linking analysis using mass spectrometry, isotope-coded cross-linkers, and integrated computational data processing. J Proteome Res. 2006; 5:2270–82. [PubMed: 16944939]

Sennels L, Bukowski-Wills JC, Rappsilber J. Improved results in proteomics by use of local and peptide-class specific false discovery rates. BMC Bioinformatics. 2009; 10:179. [PubMed: 19523214]

Sharon M, Taverner T, Ambroggio XI, Deshaies RJ, Robinson CV. Structural organization of the 19S proteasome lid: insights from MS of intact complexes. PLoS Biol. 2006; 4:e267. [PubMed: 16869714]

Singh P, Panchaud A, Goodlett DR. Chemical cross-linking and mass spectrometry as a low-resolution protein structure determination technique. Anal Chem. 2010; 82:2636–42. [PubMed: 20210330]

Singh P, Shaffer SA, Scherl A, Holman C, Pfuetzner RA, Larson Freeman TJ, Miller SI, Hernandez P, Appel RD, Goodlett DR. Characterization of protein cross-links via mass spectrometry and an open-modification search strategy. Anal Chem. 2008; 80:8799–806. [PubMed: 18947195]

Sinz A. Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions. Mass Spectrom Rev. 2006; 25:663–82. [PubMed: 16477643]

Sinz A. Investigation of protein-ligand interactions by mass spectrometry. ChemMedChem. 2007; 2:425–31. [PubMed: 17299828]

Steen H, Mann M. The ABC's (and XYZ's) of peptide sequencing. Nat Rev Mol Cell Biol. 2004; 5:699–711. [PubMed: 15340378]

Suchanek M, Radzikowska A, Thiele C. Photo-leucine and photo-methionine allow identification of protein-protein interactions in living cells. Nat Methods. 2005; 2:261–7. [PubMed: 15782218]

Sun TT, Bollen A, Kahan L, Traut RR. Topography of ribosomal proteins of the Escherichia coli 30S subunit as studied with the reversible cross-linking reagent methyl 4- mercaptobutyrimidate. Biochemistry. 1974; 13:2334–40. [PubMed: 4208778]

Swaim CL, Smith JB, Smith DL. Unexpected products from the reaction of the synthetic cross-linker 3,3′-dithiobis(sulfosuccinimidyl propionate), DTSSP with peptides. J Am Soc Mass Spectrom. 2004; 15:736–49. [PubMed: 15121203]

Syka JE, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. Proc Natl Acad Sci U S A. 2004; 101:9528–33. [PubMed: 15210983]

Tang X, Bruce JE. Chemical cross-linking for protein-protein interaction studies. Methods Mol Biol. 2009; 492:283–93. [PubMed: 19241040]

Tang X, Munske GR, Siems WF, Bruce JE. Mass spectrometry identifiable cross-linking strategy for studying protein-protein interactions. Anal Chem. 2005; 77:311–8. [PubMed: 15623310]

Trakselis MA, Alley SC, Ishmael FT. Identification and mapping of protein-protein interactions by a combination of cross-linking, cleavage, and proteomics. Bioconjug Chem. 2005; 16:741–50. [PubMed: 16029014]

Trnka MJ, Burlingame AL. Topographic studies of the GroEL-GroES chaperonin complex by chemical cross-linking using diformyl ethynylbenzene: the power of high resolution electron transfer dissociation for determination of both peptide sequences and their attachment sites. Mol Cell Proteomics. 2010; 9:2306–17. [PubMed: 20813910]

Vellucci D, Kao A, Kaake RM, Rychnovsky SD, Huang L. Selective Enrichment and Identification of Azide-tagged Cross-Linked Peptides Using Chemical Ligation and Mass Spectrometry. J Am Soc Mass Spectrom. 2010

Venselaar H, Joosten RP, Vroling B, Baakman CA, Hekkelman ML, Krieger E, Vriend G. Homology modelling and spectroscopy, a never-ending love story. Eur Biophys J. 2010; 39:551–63. [PubMed: 19718498]

Wells JM, McLuckey SA. Collision-induced dissociation (CID) of peptides and proteins. Methods Enzymol. 2005; 402:148–85. [PubMed: 16401509]

Xu H, Hsu PH, Zhang L, Tsai MD, Freitas MA. Database Search Algorithm for Identification of Intact Cross-Links in Proteins and Peptides Using Tandem Mass Spectrometry. J Proteome Res. 2010

Yan F, Che FY, Rykunov D, Nieves E, Fiser A, Weiss LM, Hogue Angeletti R. Nonprotein based enrichment method to analyze peptide cross-linking in protein complexes. Anal Chem. 2009; 81:7149–59. [PubMed: 19642656]

Yang L, Tang X, Weisbrod CR, Munske GR, Eng JK, von Haller PD, Kaiser NK, Bruce JE. A photocleavable and mass spectrometry identifiable cross-linker for protein interaction studies. Anal Chem. 2010; 82:3556–66. [PubMed: 20373789]

Young MM, Tang N, Hempel JC, Oshiro CM, Taylor EW, Kuntz ID, Gibson BW, Dollinger G. High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry. Proc Natl Acad Sci U S A. 2000; 97:5802–6. [PubMed: 10811876]

Zelter A, Hoopmann MR, Vernon R, Baker D, Maccoss MJ, Davis TN. Isotope Signatures Allow Identification of Chemically Cross-Linked Peptides by Mass Spectrometry: A Novel Method to Determine Interresidue Distances in Protein Structures through Cross-Linking. J Proteome Res. 2010

Zhang H, Tang X, Munske GR, Tolic N, Anderson GA, Bruce JE. Identification of protein-protein interactions and topologies in living cells with chemical cross-linking and mass spectrometry. Mol Cell Proteomics. 2009; 8:409–20. [PubMed: 18936057]

Zhang Q, Crosland E, Fabris D. Nested Arg-specific bifunctional crosslinkers for MS-based structural analysis of proteins and protein assemblies. Anal Chim Acta. 2008; 627:117–28. [PubMed: 18790135]

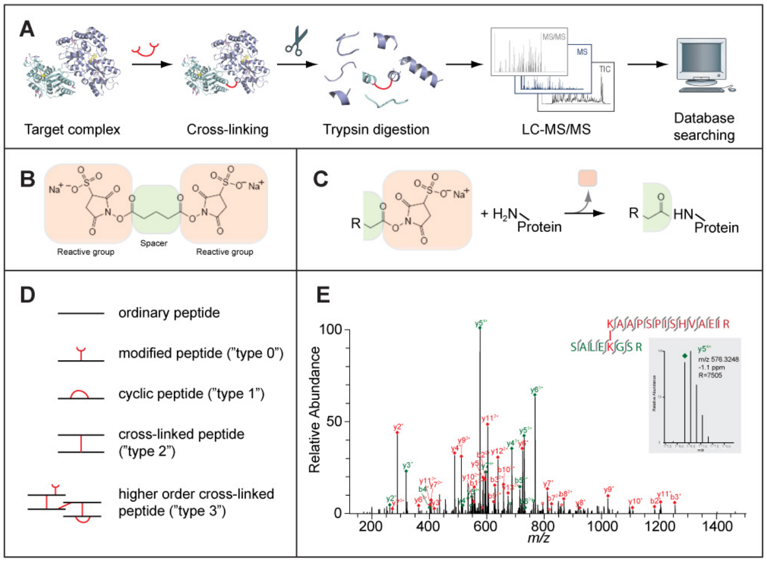

Figure 1.

(A) Outline of the cross-linking/mass spectrometry process. A target complex is cross-linked in solution and digested with trypsin into peptides. The peptides are analysed by liquid chromatography coupled high-resolution mass spectrometry (LC-MS/MS) to obtain high-resolution masses and fragment masses (high/high) for cross-linked peptides. The fragmentation spectra of all peptides are subjected to database searching to identify cross-linked peptides. As an optional step, cross-linked peptides can be enriched before their LC-MS analysis. **(B)** A typical cross-linker, here Bis[Sulfosuccinimidyl] glutarate (BS2G), is composed of two reactive groups on either end separated by a spacer. This cross-linker reacts with primary amines (lysine side chain, protein N-terminus). Others target thiols (cysteine side chain) or activate carboxylic acids (aspartate, glutamate, protein C-terminus) for reaction with primary amines. **(C)** Reaction of a cross-linker with a primary amine. Part of the cross-linker, the leaving group, is replaced by the primary amine to form a covalent bond between the spacer and the amine. In this case, a peptide bond is formed. R can stand for either the rest of the cross-linker or may contain another protein, if the cross-linker had already reacted on its other end. **(D)** Peptides types that can be observed after cross-linking and trypsin digestion. **(E)** High resolution fragmentation spectrum of a cross-linked peptide obtained on an LTQ-Orbitrap mass spectrometer (adapted from (Chen et al., 2010)). Fragment peaks are annotated in red or green, depending on the peptide that fragmented and following the naming convention for peptides (y: C-terminal fragment, b: N-terminal fragment, both as a result of dissociating the peptide bond in the peptide back bone, followed by the number of amino acids included in the fragment and the charge of the fragment). All observed fragments are also indicated as bond cleavages between amino acids in the two cross-linked peptides. In this case, virtually all possible fragments of the peptide pair have been matched and virtually all peaks have been annotated resulting in a high-confidence identification of this cross-link. The inset shows a zoom onto one fragment peak (m/z 576, 3248) which matched with -1.1 ppm to the proposed peptide sequence. The high resolution of the spectrum (R 7505 for this peak) allows clear separation of the isotope peaks and consequently assignment of the fragment's charge state.

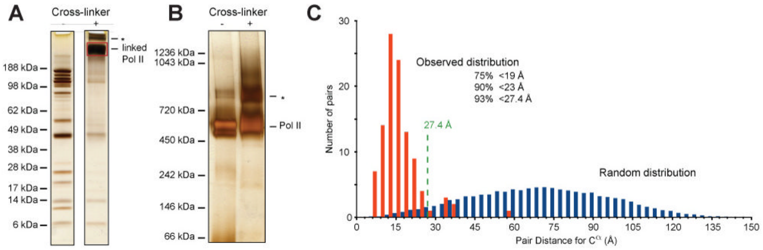

Figure 2.

Benchmarking the cross-linking/mass spectrometry process using *S. cerevisiae* RNA polymerase II (Pol II) and its crystal structure. **(A)** The subunits of Pol II are separated by denaturating gel electrophoresis (SDS PAGE) and visualized by silver staining. The individual subunits can be seen as separate bands before the addition of cross-linker (here Bis(Sulfosuccinimidyl) suberate (BS3)). After cross-linking, these individual bands disappear and a new, high-molecular weight band appears, corresponding to the cross-linked Pol II (red box). A higher molecular weight band corresponds possibly to Pol II dimers (asterisk). **(B)** Pol II migrates under native conditions mostly as a single band, both in the absence and presence of cross-linking. Under both conditions, some Pol II dimerization is observed (asterisk). **(C)** Distribution of alpha-carbon distances for lysine pairs in the crystal structure of Pol II (PDB 1WCM) (Armache et al., 2005) when scaling the distance distribution for all random lysine pairs in the crystal structure to 106 pairs (blue) and when taking the distance measure of those 106 pairs that were observed by cross-linking (red) (Chen et al., 2010). The predicted upper limit for cross-linkable lysine pairs in the crystal structure is here 27.4 Angstrom. This upper limit includes the length of lysine side chains (2 × 6.5 Angstrom), the length of the spacer (max. 11.4 Angstrom) and an estimation of the positional error in the crystal structure (1.5 Angstrom for surface residues). The upper limit does not consider the possibility of conformation changes or vibrations of the complex in solution. The observed distribution of cross-linked pairs is clearly not random and fulfils largely the theoretically predicted distance threshold for cross-linkable pairs. (all adapted from (Chen et al., 2010))

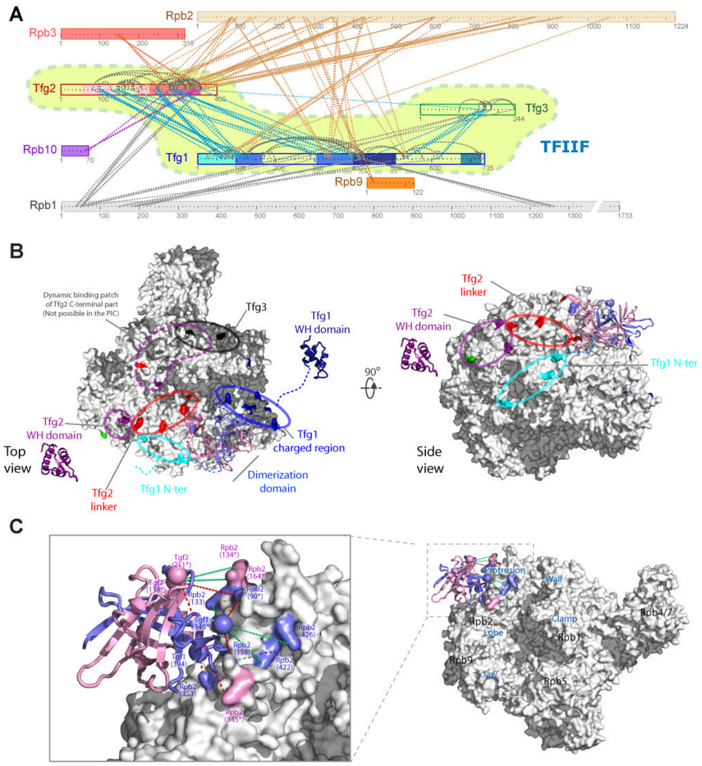

Figure 3.

Cross-linking/mass spectrometry analysis of *S. cerevisiae* RNA polymerase II (Pol II) bound to transcription factor IIF (TFIIF). **(A)** Linkage map showing the sequence position of all observed cross-linked residue pairs within TFIIF and between TFIIF and Pol II. Connections between residues are blue within TFIIF or colour coded by Pol II subunit for cross-links between Pol II and TFIIF. Sequence regions of TFIIF subunits are colour coded (Tfg1: N-terminal tail, 2 × dimerization domain, charged region, winged-helix (WH) domain; Tfg2: 2 × dimerization domain, linker, WH domain). **(B)** Residues of Pol II colour coded by region in TFIIF subunits they cross-link with. **(C)** Homology model of the Tfg1-Tfg2 dimerization domain positioned on the Pol II structure (PDB 1WCM) with cross-linked residues labelled by proteins and residue number. Dashed lines connect pairs of residues that were used for the positioning, either because they were observed to cross-link or because they are the closest residues in the structure (denoted by an asterisk behind their residue number). (all adapted from (Chen et al., 2010))