Protein Structure Determination Using a Combination Of Cross-linking, Mass Spectrometry and Molecular Modeling

Dmitri Mouradov[1], Gordon King[2], Ian L. Ross[2,3], Jade K. Forwood[1], David A. Hume[1,2,3], Andrea Sinz[4], Jennifer L. Martin[1,2], Bostjan Kobe[1,2] and Thomas Huber[1]

[1] School of Molecular and Microbial Sciences;. [2] Institute for Molecular Bioscience and ARC Special Research Centre for Functional and Applied Genomics;. [3] CRC for Chronic Inflammatory Diseases, The University of Queensland, Brisbane, Queensland, Australia; [4] Institute of Analytical Chemistry, Biotechnological-Biomedical Center, University of Leipzig, D-04103 Leipzig, Germany.

Corresponding author: Bostjan Kobe, School of Molecular and Microbial Sciences, The University of Queensland, Brisbane, Queensland 4072, Australia; email: b.kobe@uq,edu.au; phone: +617-3365-2132; fax: +617-3365-4699.

Running title: Structure determination by cross-linking

Abstract

Cross-linking in combination with mass spectrometry can be used as a tool for structural modeling of protein complexes and multi-domain proteins. While cross-links represent only weak structural constraints, the combination of a limited set of experimental cross-links with molecular docking/modelling is often sufficient to determine the structure of a protein complex or multi domain protein at low resolution.

Key Words: cross-linking, mass spectrometry, molecular modellibg, structure determination

1. Introduction

Most current structural genomics initiatives have focused on high-throughput structure determination of individual proteins using x-ray crystallography and NMR spectroscopy. As our understanding of protein folds becomes more comprehensive the next step of proteomics will have to include multi-protein and transient complexes that are essential for many cellular functions. Greater understanding of protein-protein interactions will lead to a deeper understanding of the regulation of cellular processes and to significant benefits to biotechnology.

Even though some high resolution initiatives focus on protein complexes, their success rates are low, judging by deposits in the Protein Data Bank (PDB). These low success rates are associated with the difficulty in crystallizing complexes. An emerging approach is to derive a set of sparse distance constraints using chemical cross-linkers, to map out residues in the protein interaction interface, and then assemble relevant orientation of protein complex structure by computational means. This approach is particularly powerful when partial structural information is available, for example, the structures of individual proteins in a complex or individual domains in a multi-domain protein. In such cases, this technique allows high-throughput low-resolution structure determination even with a limited number of approximate distance constraints.

Chemical cross-linkers have been successfully used for many years to derive low resolution structural information of proteins. Such cross-linking experiments gave insight into interaction modes of large protein complexes as well as identifying interacting proteins. However, the demonstration of the presence of an interaction does not include information about how the proteins dock with respect to each other.

Recent advances in mass spectrometry (MS) (*1*) have allowed for analysis of a large number of enzymatically digested peptides (*2*) and hence identification of the exact insertion points of low-abundance cross-links (*3-6*) and have opened up a new perspective on the use of cross-linkers in combination with computational structure

prediction (*7, 8*). The approach uses chemical cross-linking information with molecular docking, so that the cross-links are treated as explicit constraints in the calculations. By using a simple rigid body docking algorithm and a small number of constraints it is possible to narrow down the conformation of a protein complex to only a few configurations and hence identify the interface of interaction.

1.1. Cross-linkers

Many types of cross-linkers are commercially available, offering a wide variety of probing tools. Cross-linking reagents can be divided into four general classes: homobifunctional, heterobifunctional, trifunctional, and zero–length (*9*).

Homobifunctional cross-linkers contain a spacer (of various lengths) linking two identical functional groups (**Fig. 1A**). These cross-linkers are designed for simple one-step reactions. Heterobifunctional cross-linkers contain two different functional groups linked by a spacer (**Fig. 1B**). Having two different reactive groups allows greater control over the cross-linking reaction enabling the incorporation of each reactive group in separate steps. For example, a NHS ester reactive group can be conjugated with amine groups on one protein, while a second photo-reactive group will be brought to reaction once the binding partner has been added.

In the case of trifunctional reagents (**Fig. 1C**) the third functional group can either incorporate an affinity tag (*10*) or another reactive group to cross-link a third site. The presence of an affinity tag, such as biotin, allows a much simpler isolation of inserted cross-linked peptides after enzymatic digestion. Sulfo-SBED is a trifunctional cross-linker that has been used in mapping protein interfaces (*11*).

Zero-length cross-linkers (*12*) covalently link two proteins without the incorporation of a linker (**Fig. 1D**).

1.2. Functional Groups

With cross-linking becoming widely used for numerous purposes, hundreds of cross-linking reagents have become commercially available, however, their reactions with proteins are based on a limited number of organic reactions. A large variety of cross-linking reagents are created by mixing and matching these functional groups and varying the spacer lengths between them. The most commonly used functional groups will be described here. While most functional groups are highly specific it must be noted that non-specific reactions may also occur at various physiological conditions.

1.2.1. Amine-reactive Cross-linking Reagents (NHS esters, imidoesters)

Amine-reactive cross-linking reagents are among the first sidechain-specific cross-linking reagents to be developed. The chemistry of amine-reactive cross-linkers is based on reaction with primary amines. Two main amine-reactive functional groups are used; NHS esters and imidoesters.

NHS esters react with ε-amines on lysine residues and free α-amines of N-termini from proteins, forming amide bonds (**Fig. 2A**) (*13, 14*). The reaction can be followed by monitoring NHS release by UV spectroscopy at 260 nm. Both water-soluble and water-insoluble reagents are available, with the water-soluble counterpart containing a sulfonate group. Reactions are usually performed at temperatures between 4 ^{o}C and room temperature at a pH between 7 and 9. The reaction is usually quenched with primary amine-containing buffers such as Tris (*7*). NHS esters might also react with serines and tyrosines (*15*).

A disadvantage of amine reactive cross-linkers is that a reacted lysine is cleaved with a much lower frequency by trypsin (the most widely used proteolytic enzyme in this field). This results in larger fragments after digestion that are more difficult to identify by mass spectrometry.

Imidoesters also react with ε-amines and free α-amines of N-termini, but form an amidine linkage. Imidoesters react very rapidly with amines at alkaline pH; however, they also have a short half-life for hydrolysis (*16*). It has been shown that the reaction

forming the amidine bond is reversible at high pH (*17*). Diimidoesters were used to cross-link neighboring proteins in the 30S ribosomal subunit, identifying interactions between S7-S9 and S13-S19 subunits (*18*).

1.2.2. Maleimides (Sulfhydryl-reactive reagents)

Maleimides target thiol groups of cysteines forming thioether bonds (**Fig. 2B**). Such cross-linkers have been successfully utilized to map interaction sites (*19*). There are many advantages of using maleimide groups in cross-linking including high yield of cleavable products and the ability to perform sequential coupling. At pH 6.5-7.5 maleimides react specifically with sulfhydryl groups to form irreversible bonds (*20*), however at more basic pH a reaction may occur with amines. An advantage of using maleimides over amine reactive groups is that the reaction leaves lysines accessible for tryptic digestion. Sulfhydryls can be introduced into proteins by reacting primary amines with 2-iminothiolane as the number of free thiols in most proteins is low.

1.2.3. Photo-reactive Reagents

Azides and other photoreactive groups become reactive in the presence of UV or visible light. The radical reaction at the presence of light forms a nitrene group that can be inserted into C-H (**Fig. 2C**) or N-H. Heterobifunctional and trifunctional cross-linkers containing a photoreactive group allow a simple two-step reaction where one reactive site of the cross-linker is chemically conjugated in the absence of light, whereas the photo-reactive site is conjugated to the protein after addition of the binding partner. This minimizes the potential for self-reactivity and increases the probability that any of the cross-linked partners is present as a specific heterodimer. Another advantage is the non-specific reactivity of the photoreactive site allows a variety of residues to be a good candidate for cross-linking studies, by contrast to cross-linkers that require a specific amino acid for reaction. The non-specific reactivity becomes a disadvantage for identification of the exact points of cross-linking because the range of possible reaction sites is much larger. Another disadvantage of photo-reactive cross-linking reagents is that the radical reaction might cause random insertions into the protein. Hence, all species are

present in very low abundance, which greatly complicates the analysis of the resulting fragments.

1.3. Cross-linked Products

The insertion of a chemical cross-linking reagent into proteins can result in a multitude of products. It is important to understand the various types of cross-linked products in order to successfully identify them. Upon proteolysis one of three types of cross-linking products (**Fig. 3**) can be formed, these have been designated type 0, 1 or 2 (*21*). Type 0 (or a "dead-end") product results from a cross-linker that has reacted to a peptide at one of its reactive sites, but is hydrolyzed at the other end, thus, losing its ability to form a cross-link. This type of product provides information on the surface accessibility of the amino acid. A type 1 (or "cyclic") product results when both reactive groups cross-link within the same peptide following trypsin digestion. Intra-peptide products provide limited distance information as cross-linked amino acids will generally be close in primary sequence. A type 2 or inter-peptide cross-linked product results from cross-linking of two residues that are either well-separated in primary sequence, with cleavage sites between them, or on separate protein molecules. These products usually provide the most useful set of distance constraints.

1.4. Identification of Cross-links

The most significant challenge in structural studies using cross-linkers is the actual identification of cross-linked peptides in the mass spectrum. The challenges of analyzing cross-linked products arise from the relatively low abundance of cross-linked material in comparison to underivatized peptides, which leads to difficulties in assignment, especially if no sequence data (MSMS) are available. Assignment by mass alone is problematic in cases when masses can be assigned to native peptides as well as different cross-linked products. Tandem MS (MS/MS) is also challenging as complex spectra are formed when two cross-linked peptides are fragmented. Over the past few years, various approaches have been devised to address this problem. Innovations devised to address this problem include isotope labeling of cross-linkers, introduction of cleavable cross-

linkers and addition of affinity tags to cross-linkers (**9**). These methods provide different approaches to facilitate correct identification of cross-linked species.

1.4.1. Isotope Labeling

Isotope labeling is a common approach in mass spectrometry. In this case, the use of a mixture of "light" and "heavy" cross-linking reagents allows for easy visual identification of cross-link doublets in a spectrum using standard mass spectrometric ionization techniques, such as MALDI-TOF (matrix-assisted laser desorption/ionization-time of flight) or ESI (electrospray ionization). Isotope labeling involves the use of a 1:1 mixture of identical cross-linking agents differing for example, only by the number of deuterium atoms of their chemical composition. Cross-linking agents, such as BS^3-d_0 and BS^3-d_4 which differ by the presence of 4 deuterium atoms are commercially available (**22**). The assumption is that the chemical reactivities of the isotope labeled reagents are identical, thereby yielding doublets separated by the defined mass difference (e.g. 4.025 amu) in the mass spectrum. This usually enables unambiguous discrimination between underivatized peptide peaks and those modified by the specific incorporation of the synthetic cross-linker.

1.4.2. Cleavable Cross-linking Reagents

Since fragmentation spectra provide much more reliable peptide identifications than the parent mass alone, methods to enable cross-linker cleavage (thereby reconstituting linear peptides with interpretable fragmentation spectra) are desirable. This can be achieved by the insertion of a disulfide bond or carboxylic ester into the cross-linker. One such cross-linker is DTSSP (3,3´-Dithiobis[sulfosuccinimidylpropionate]), which contains a thiol cleavable bond. Comparing the mass spectra of unreduced and reduced peptide allows for quick identification of putative cross-linked fragments (**23, 24**). Masses corresponding to peptides that disappear after reduction are listed as putative cross-linked peptides, and further identification can be carried out by finding the two halves of the cross-linked peptides in the reduced sample.

1.4.3. Affinity-tagged Cross-linkers

An obvious solution to the problem of low abundance cross-links in a peptide digest is to enrich or purify only those peptides containing the cross-link. Trifunctional cross-linkers containing an affinity tag are powerful tools for cross-linker identification and assignment. With a simple purification step (avidin column in the case of a biotin-tagged cross-linker) non-cross-linked native peptides will be eluted, leaving only peptides with bound cross-linker for mass spectrometric analysis. This allows for low abundant cross-linked peptides to be concentrated without native peptides dominating the resulting spectra (*11*).

1.4.4. MALDI-TOF/TOF-MS of Cleavable Cross-linkers

At the forefront of cross-link identification methods a new promising protocol is being developed combining the DTSSP cleavable cross-linker and MALDI TOF/TOF-MS. This breakthrough uses the characteristic asymmetric breaking of disulfide bonds. A distinct 66 amu-doublet is observed after fragmentation of the disulfide bond from a DTSSP cross-linker (**Fig. 4**). The asymmetric fragmentation of the DTSSP cross-linker gives rise to dehydroalanine and thiocysteine analogues which are 54 and 120 mass units larger then the parent peptides.

1.5. Mass Spectrometry

Various mass spectrometric methods are available for identification of cross-linked species. Each technique has its own advantages so various techniques are sometimes required for a thorough analysis. A liquid chromatography (LC) separation step before mass spectrometry has become an invaluable tool as it reduces ion suppression.

1.5.1. MALDI-TOF Mass Spectrometry

One of the MS analyses available to identify possible cross-linked peptides is MALDI-TOF mass spectrometry and has been utilized in numerous studies (*3, 10, 25, 26*). By comparing modified and non-modified digests one can assign potential cross-linked peptides by mass alone (assuming a well-calibrated instrument). MALDI-TOF mass spectra allow for a quick identification of putative cross-links after overlaying the

spectra. Another advantage of this method is the low amount of material needed to carry out the analysis. The biggest disadvantage of using MALDI-MS is the preference for arginine-containing sequences as opposed to lysine-containing sequences. Lysine-containing peptides are often suppressed and their signal is indistinguishable from background noise whereas arginine has a greater ionization efficiency derived from its more basic side chain (*27*). This problem of preference can be overcome by the use of an alternative protease to trypsin or by converting lysines to homo-arginines using O-methylisourea (*28*).

Recently MALDI-LC/MS/MS using a MALDI-TOF/TOF mass spectrometer has been proven to be an invaluable tool in identifying DTSSP cross-linked peptides (or other disulfide-based cleavable cross-linkers). This type of mass spectrometer is valuable for identifying all types of cross-linkers using a preceding LC separation step and MS/MS sequencing. However, it is the distinct fragmentation pattern for DTSSP cross-links that makes this cross-linker such an invaluable tool. The downside is that this method has the potential of generating a large number of spectra that must be examined, and no automated tools are yet available.

1.5.2. Electrospray Ionisation

Electrospray ionisation (ESI) mass spectrometers are also widely used for analysis of cross-linked peptides (*7*). With a separation technique such as a liquid chromatography (LC) system attached, ESI-MS presents a powerful tool for the identification of cross-linked peptides. Another advantage of ESI is the lack of preference for specific residues which is prevalent in MALDI-TOF MS.

1.5.3. FTICR

Fourier transform ion cyclotron resonance (FTICR) mass spectrometers have also been utilized for cross-linking studies (*4*). FTICR offers both ultra-high resolution and a gas-phase purification step. Its ultra-high accuracy allows for unambiguous identification of cross-linked species by mass alone (*29*).

1.6. Software

Assigning mass spectrometry mass lists to cross-linked peptides can be done via various internet tools such as ASAP (Automatic Spectrum Assignment Program), MS2Assign (*21*) and MS2PRO (all three can be found at http://roswell.ca.sandia.gov/~mmyoung/index.html). Such tools allow for analysis of various cross-linker reagents, proteases and MS equipment. MS2PRO and MS2Assign can also be used to assign MS/MS fragmentation of underivatized and cross-linked peptides. In our lab, custom software was developed to help identify cross-links from mass spectra.

1.7. General Methodology

Two general approaches can be applied to cross-linking studies: top-down and bottom-up (*9*). The top-down approach analyzes intact cross-linked proteins, where gas phase purification and fragmentation is carried out within the mass spectrometer. The more widely used bottom-up approach involves separation of cross-linked species by size exclusion, proteolytic digestion and mass spectrometry (**Fig. 5**).

The introduction of cross-links into protein complexes can be carried out in a one- or two-step reaction depending on the chosen cross-linking reagent. Homobifunctional cross-linkers are mostly employed in a one-step reaction where both functional groups react with the protein complex. Using heterobifunctional reagents can allow for a two-step reaction where one of the functional groups is reacted with one of the proteins in the complex, the second protein is added and the cross-linking conditions are changed to allow the other functional group to react. This process increases the chance for formation of inter-protein cross-links.

Here we provide a protocol for the bottom-up approach using homobifunctional cross-linking reagents. The protein complex of interest is mixed with a cross-linking mixture and allowed to react. After quenching, the resulting cross-linked and non-cross-linked proteins are run on SDS-PAGE for separation. The band of interest is excised and used

for *in-gel* digestion. The resulting peptides are submitted to mass spectrometric anlysis and the resulting masses are analyzed by using software packages to identify cross-linked products. Commputational modeling and/or docking is carried out based on distance constraints from identified cross-links.

2. Materials

2.1. Chemical Cross-linking

1. Bis(sulfosuccinimidyl) suberate cross-linker (Sigma, S5799)). Store at -80 $^\circ$C.

2. DTSSP cross-linker (Pierce, 21578) Store at -80 $^\circ$C.

3. 100 mM Tris-HCl pH 8.0.

4. Protein buffer (primary amine free buffer with pH between 7 – 9).

2.2. SDS-PAGE

1. Pre-cast 4-20% SDS-PAGE gel (non reducing).

2. 5X running buffer: 125 mM Tris, 960 mM glycine, 0.5% (w/v) SDS. Store at room temperature.

3. Prestained molecular weight markers (Bio-Rad).

4. Coomassie-blue stain: 0.5 g Coomassie Blue (0.1% w/v in final solution, 50 ml acetic acid (10% v/v in final solution), 100 ml methanol (20% v/v in final solution), 350 ml water.

5. Destain buffer: 400 μL ethanol, 100 μL acetic acid, 500 μL water.

2.3. In-gel Digestion and Extraction

1. Wash buffer: 50% CH_3CN, 50 mM NH_4HCO_3.

2. 0.5 mg/mL trypsin (Sigma). Store at -80 $^\circ$C.

3. 50 mM NH_4HCO_3.

4. Extraction buffer: 60% CH_3CN/0.1% TFA.

2.4. MALDI-MS

1. CHCA (α-cyano-4-hydroxycinnamic acid) matrix (Agilent Technologies #G2037A)).

2. ProteoMass peptide and protein MALDI-MS calibration mix (Sigma # MSCAL1-1KT).

3. 1 M DTT stock.

2.5. LC/MS

1. C18 capillary column (Agilent).

2. 40% CH_3CN/0.1% acetic acid.

3. 100% H_2O/0.1% acetic acid.

4. 100% CH_3CN/0.1% acetic acid.

2.6. Peak Assignment

1. Preferred peak assignment program capable of assigning cross-linked peptides (http://roswell.ca.sandia.gov/~mmyoung/index.html).

2.7. Modeling

1. Structure of each protein/domain in Protein Data Bank format.

2. Rigid body docking algorithm capable of accepting distance constraints between atoms as an input.

3. Methods

Here we describe the method used to determine the structure of latexin:carboxypeptidase A (CPA) complex (Mouradov *et al.,* 2006) and mouse acyl-CoA thioesterase (using BS^3 and DTSSP cross-linkers, and MALDI-TOF and ESI mass spectrometry).

3.1. Cross-linking

1. Combine 100 μl complex (~3 mg/mL at pH 7-9) with 10 X molar excess of cross-linking solution BS^3 (Sigma, S5799) or DTSSP cross-linkers (Pierce, 21578) in separate Eppendorf cups.

2. Incubate for 2 hours at room temperature (*see* **Note 1**).

3. Quench reaction using 100 mM Tris buffer (pH 8).

3.2 SDS-PAGE and In-gel digestion and extraction

1. Prepare running buffer by diluting 100 mL of 5 X with 400 mL of water.

2. Add running buffer to upper and lower chambers of gel unit. Load markers and 40 μL of sample in multiple lanes onto a Gradipore precast SDS-PAGE gel.

3. Separate out cross-linked complex from higher molecular aggregates by running for approximately 1.5 hours at 170 V.

4. Stain with Coomassie Brilliant Blue G250 or R250 and destain to visualize the bands (**Fig. 6**).

5. Excise the bands corresponding to your complex. You may pool 2 or more excised bands into one Eppendorf cup. Excise at least 2 bands for best results.

6. Wash the excised bands several times with of 200 μL of 50% CH$_3$CN, 50 mM NH$_4$HCO$_3$. Then dry and incubate in 5 μL of 0.5 mg/mL trypsin (Sigma) and 200 μL of 50 mM NH$_4$HCO$_3$ at 37°C overnight for digestion.

7. Transfer the digestion into clean Eppendorf cup. Perform three peptide extractions by adding 200 μL of 60% CH$_3$CN/0.1% TFA to the gels, shaking at 200 rpm for 30 min, centrifuge at 1000 g for 1 min and then pool.

8. Concentrate pooled sample using a SpeedVac or Millipore ZipTips C18 (*see* **Note 2**) system and resuspend in ~60 μL 60% CH$_3$CN/0.1% TFA.

3.3. MALDI Mass Spectrometry

1. Separate the sample cross-linked with DTSSP in two parts. One of the two samples is reduced with 10 mM DTT for 5-10 min.

2. Combine your digested sample in a 1:2 ratio with the CHCA matrix (Sigma) (*see* **Note 3**).

3. Prepare 1 μL of analyte / matrix mixture onto 96-well plate and allow to fully dry.

4. Prepare 1 μL of your low molecular weight calibration mixture (Sigma) onto the 96-well plate and allow to dry.

5. Load the sample plate and apply calibration. Each samples is initially analyzed by adding ~500 laser shots to one mass spectrum. Export the peak list (**Fig. 7**). For best results de-isotope the series first.

3.4. LC/MS

1. Separated the sample first by reverse phase HPLC (Agilent) using a C18 capillary column. The elution gradient is 0-60% (v/v) acetonitrile in 0.1 acetic acid over 45 min with a flow gradient of 0.1 μL/min. The ES spectrum is recorded on a Applied Biosystems QSTAR Pulsar mass spectrometer.

2. Manually go though spectra and extract masses (*see* **Note 4**).

3.5. Peptide Assignment

1. Analyse the set of *m/z* peaks obtained from the mass spectra using an in-house program that assigns *m/z* values to possible cross-linked peptide fragments from amino acid sequences. (assignment programs are available on the World Wide Web). (*see* **Notes 5 and 6**)

2. Identify putative cross-linked peptides by comparison of reduced and non-reduced DTSSP spectra.

3. Cross-check putatively assigned cross-linked fragments with the original spectra for validation of 'real' peaks by identifying multiple charged states (**see Note 7**).

3.6. Docking with Distance Constraints

1. Our in house docking algorithm uses a systematic six-dimensional search over all rotations in steps of 5 degrees and all cartesian translations of 1.0 Å up to ±66 Å along each coordinate axis. The distance constraint between Cα- Cα distance of cross-linked lysine residues is estimated at 25 Å for the BS³ and

DTSSP cross-linkers. A linear scaling grid cell algorithm with geometric hashing is used to check for any inter-molecular residue pairs in close spatial proximity and thus to exclude those models with steric overlap.

2. Score models by a simple hydrophobic energy score that counts the number of contacts (<8 Å) between hydrophobic amino acids. Consider the top 1000 structures.

3. Cluster the top 1000 structures based on the root mean square deviations (RMSD) of the coordinates of Cα atoms. The group with the highest average hydrophobic contacts is considered. The best docked structure is considered to be the highest scoring model based on the scoring function in that group (**Fig. 8, 9**).

4. Notes

1. Cross-linking for 2 hours may cause covalently linked higher order polymers of complexes that tend to naturally form polymers. In order to reduce formation of higher order polymers the reaction must be quenched earlier. A time-course with various quenching times is the best way to find the optimal reaction time.

2. ZipTips (C18 for peptides, C4 for proteins) are preferred for concentrating samples in preparation for mass spectrometry as it de-salts your sample. Salt interferes with both MALDI and ESI mass spectrometry.

3. MALDI-TOF mass spectra can be improved through a process of lysine guanidination (*28*). MALDI-TOF mass spectra of tryptic digests are dominated by arginine-containing peptides, while lysine terminated peptides givesignals with lower intensities. By using *O*-methylisourea you can convert lysine to homoarginine preventing this problem.

4. While software is available to extract masses from ESI files, it is strongly recommended to go through the spectra manually. Most automated software will extract many false positive masses.

5. A maximum allowed error must be defined between the observed masses and expected masses of cross-linked peptides. The error will depend on mass accuracy of the mass spectrometer that is being used. As cross-linked peptides will be in a much lower abundance than native peptides; their peak intensity will be considerably lower and might therefore have a greater error..

6. Assigning mass to peaks of cross-linked proteins can be done via various available internet tools such as ASAP (Automatic Spectrum Assignment Program). For our purposes we developed our own tool to meet our requirements.

7. Without sequencing data, cross-link assignment can be a challenging exercise. With mass-only assignments it comes down to how confident one is of the assignment. This involves checking the potential cross-linked mass against native peptide masses as well as trypsin self-cleavage fragments.

Tandem MS (MS/MS) analysis can be added as an extra step. However, fragmentation of two cross-linked peptides creates very complex spectra. New techniques are currently being perfected using cleavable cross-linkers to add another degree of certainty to the assignment as well as being able to create simpler fragmentation patterns using MS/MS.

Figure legends

Fig. 1. Examples of commercially available cross-linkers. (A) DTSSP is a homobifunctional amine reactive cross-linker. (B) SADP is a heterobifunctional cross-linker with amine and sulfhydryl-reactive groups. (C) Sulfo-SBED is a trifunctional cross-linker containing a amine reactive and photo-reactive group as well as a biotin tag. (D) EDC is a zero length cross-linker, which reacts with amine and carboxylic acid groups.

Fig. 2. Many chemical reactions are possible between cross-linkers and proteins. (A) Reaction of NHS-ester cross-linker and an amine forming an amide bond. (B) Reaction of maleimide and a sulfhydryl forming a thioether bond. (C) One of many possible radical reactions of phenyl azide.

Fig. 3. The three main types of cross-linking products formed by bifunctional cross-linking reagents.

Fig. 4. Unique fragmentation of a DTSSP cross-linked fragments using MALDI TOF/TOF-MS. The two sets of peaks, which are 66 amu apart from each other, represent the asymmetric fragmentation of each half of the cross-linked species.

Fig. 5. The overall process for determining protein complex determination using the BS^3 or DTSSP cross-linker and mass spectrometry. (A) The addition of BS^3 or DTSSP (homobifunctional lysine specific cross-linker) to the complex. (B) The cross-linked complex is separated on an SDS-PAGE gel for in-gel digestion and extraction of peptides. (C) The extracted peptides were analysed by MALDI-TOF-MS and LC/ESI-MS. Form the mass spectra a list of masses is extracted. (D) These masses are then assigned using an in-house program that searches for possible cross-links and these cross-links are used as distance constraints for rigid body docking. (E) The best structure is determined using a scoring function based on hydrophobic interactions.

Fig. 6. SDS-PAGE of cross-linked and non-cross-linked samples. (A) Cross-linking carboxypeptidase A and latexin shows the formation of a 1-to-1 complex. (B) Cross-linking of murine acyl-CoA thioesterase shows the formation of a cross-linked trimer.

Fig. 7. A MALDI-TOF mass spectrum from cross-linked and non-crosslinked tryptically digested murine acyl-CoA thioesterase. The arrow show a mass linked with a potential cross-linked fragment. These assignments are made based on mass alone and must be checked against a non-cross-linked control.

Fig. 8. Docking of carboxypeptidase A (dark grey) to latexin (black) using in-house rigid body docking software. (A) The top 1000 structures are shown. (B) The top scoring cluster based on average hydrophobic interaction scoring. (C) The top scoring docked structure superimposed onto the crystal structure configuration (light grey). The crystal structure of this complex became available (**30**) shortly after this model was built, which allowed a direct comparison.

Fig. 9. Close-up of the identified interface between latexin (light) and carboxypeptidase A (dark) as predicted via docking. The dashed lines represent cross-links and the black regions represent location of cross-linked residues.
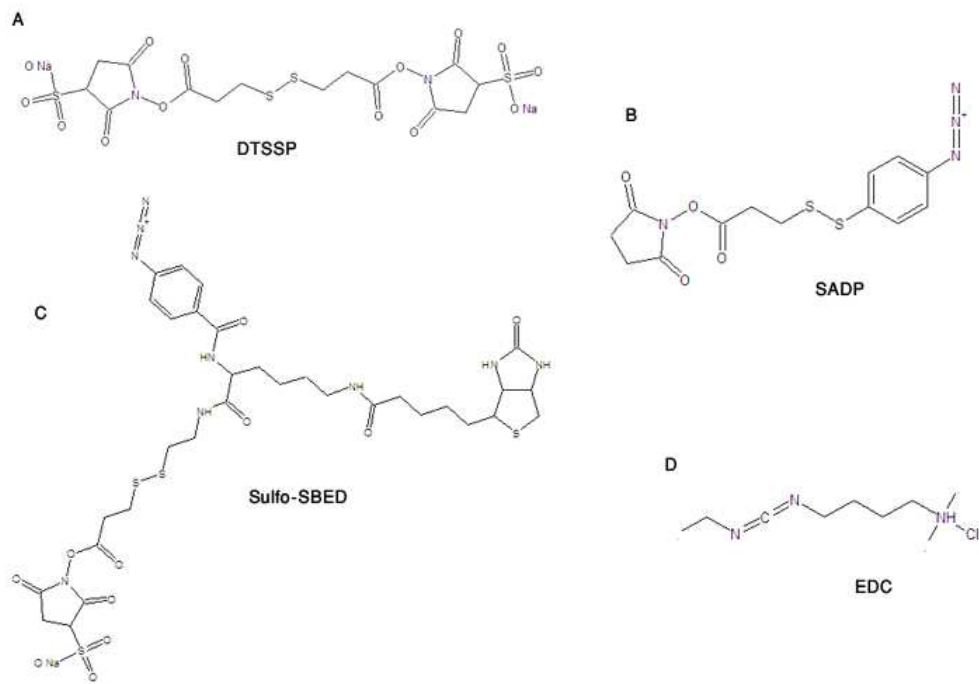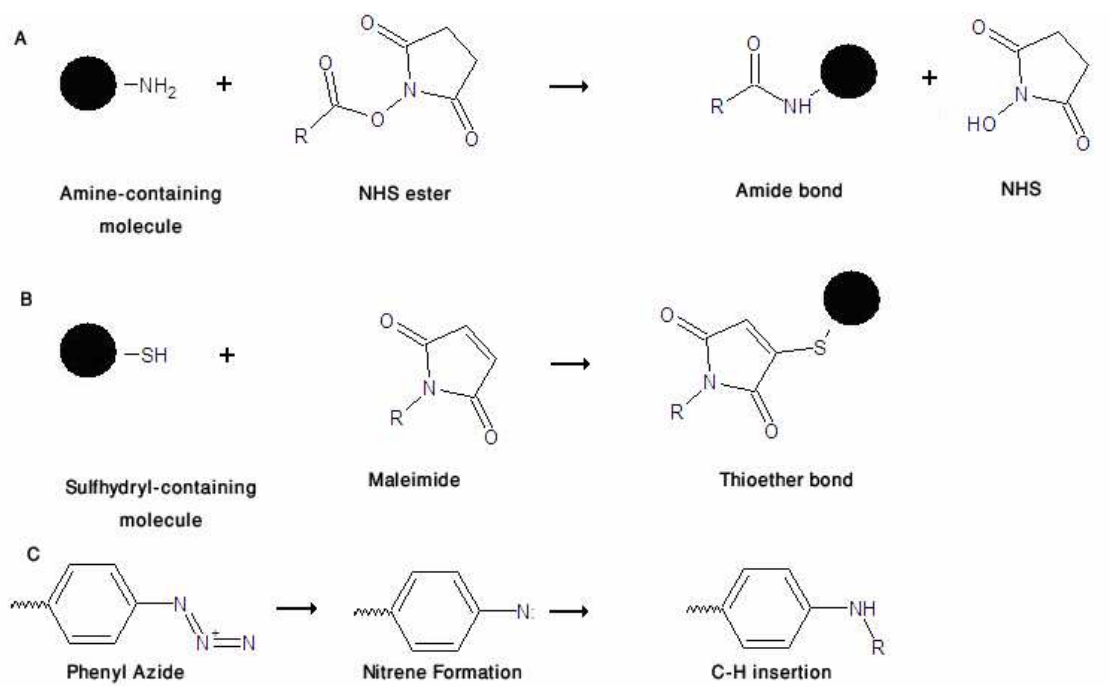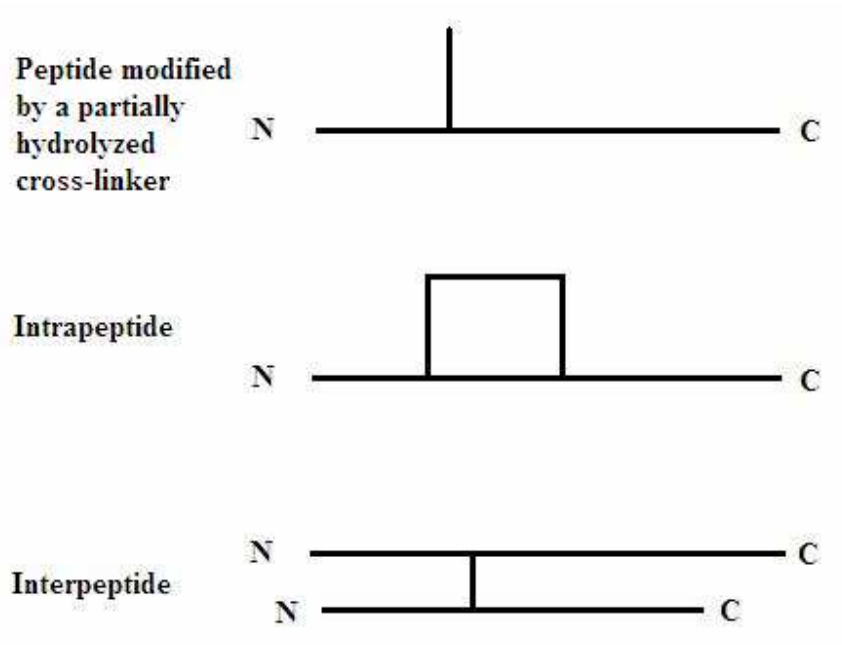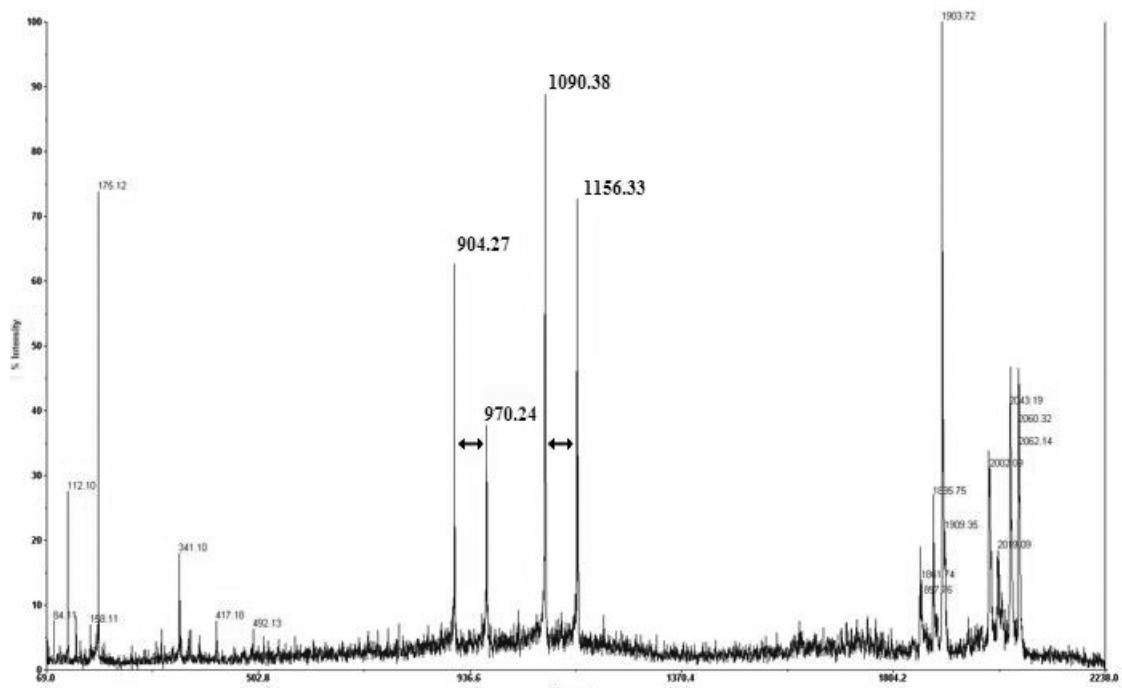
Figure 1

Figure 2

Figure 3

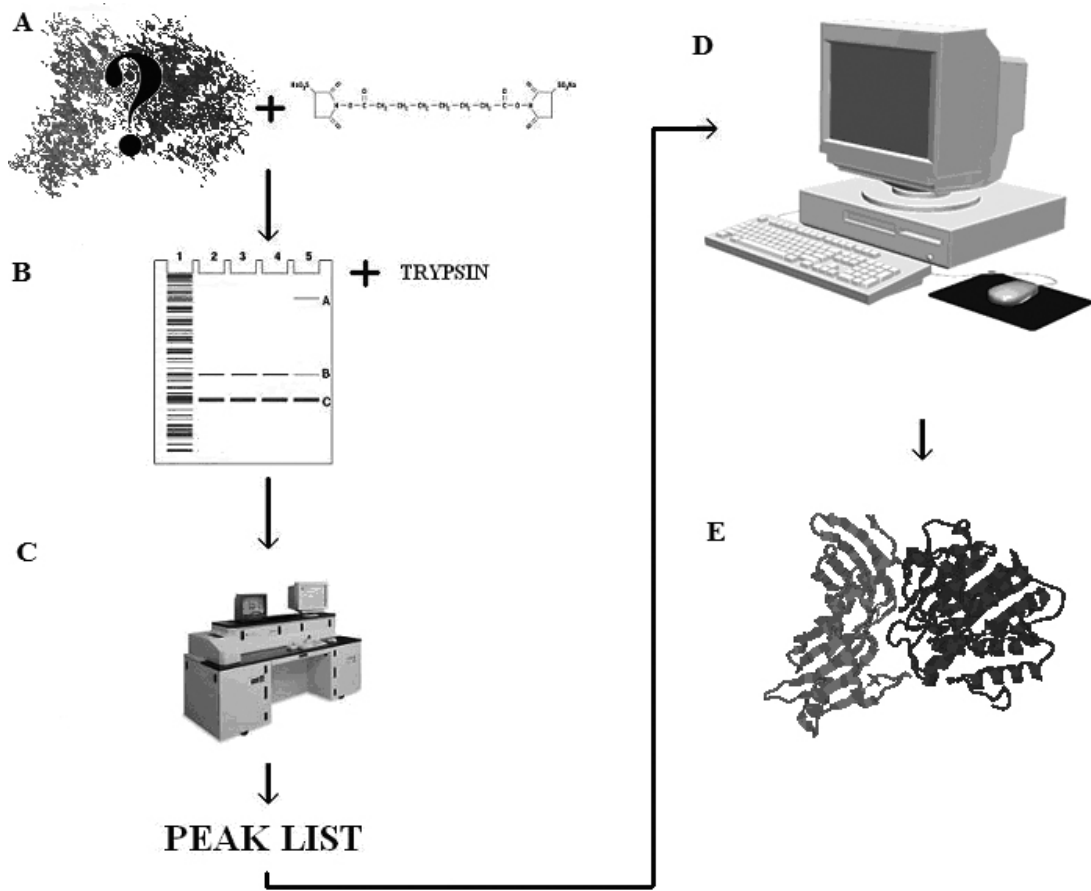Figure 4

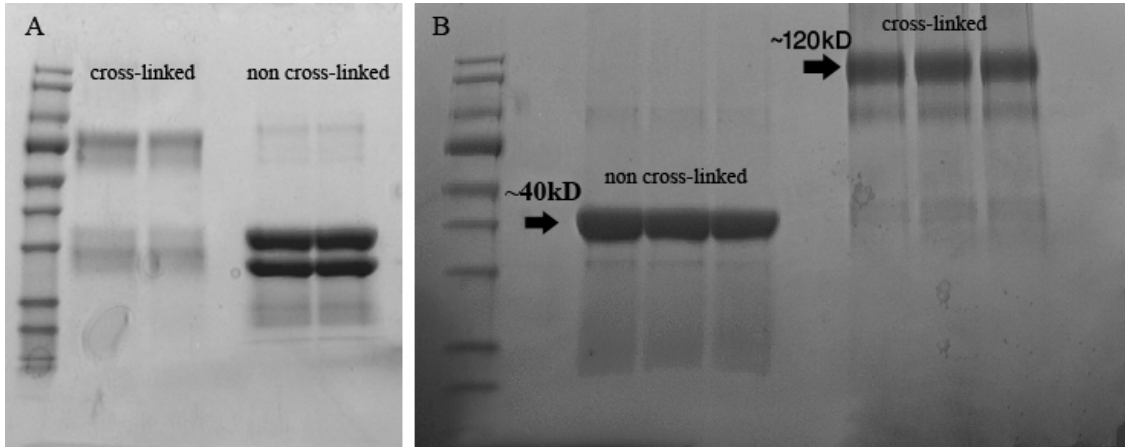Figure   5
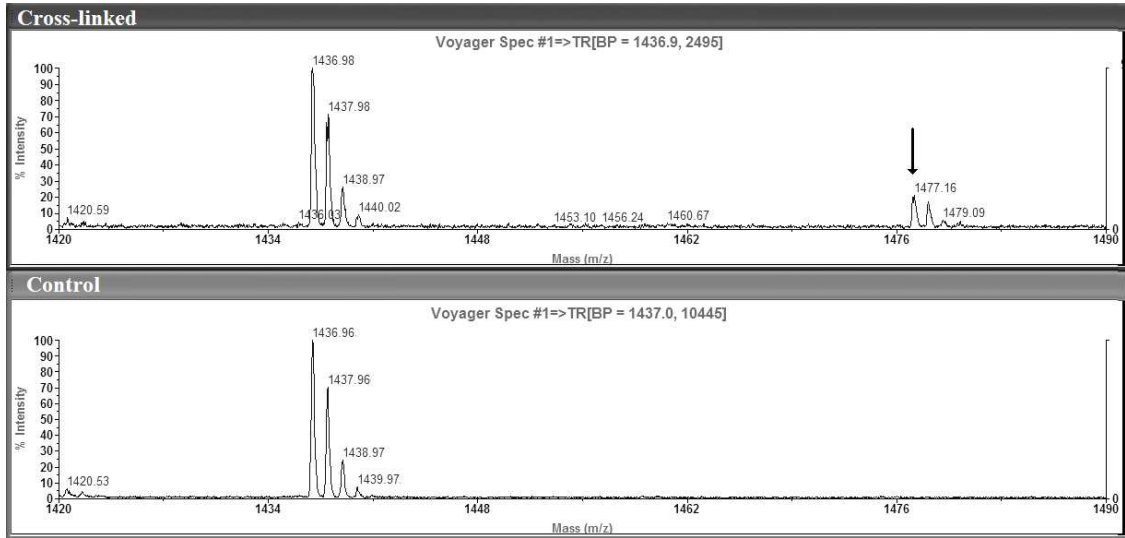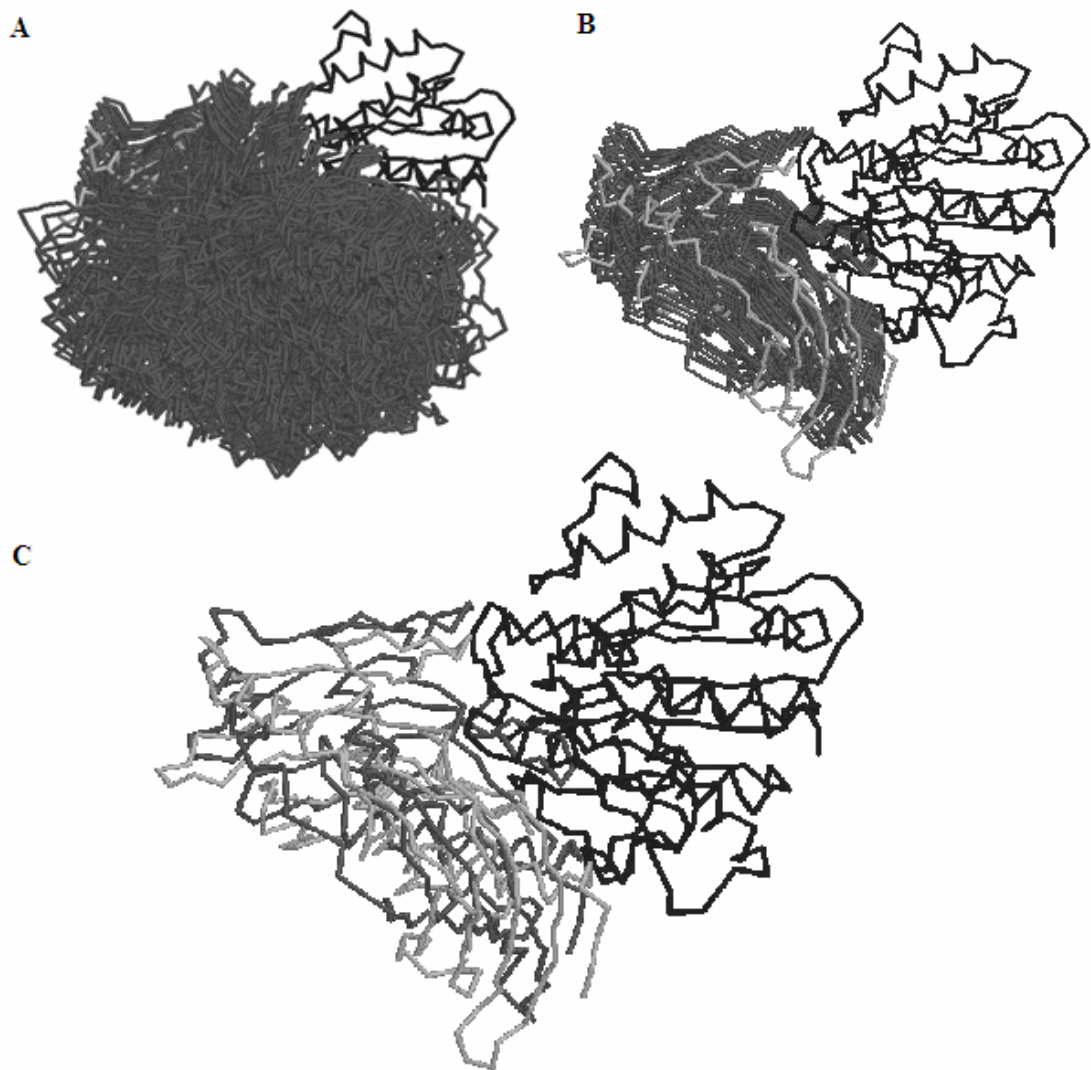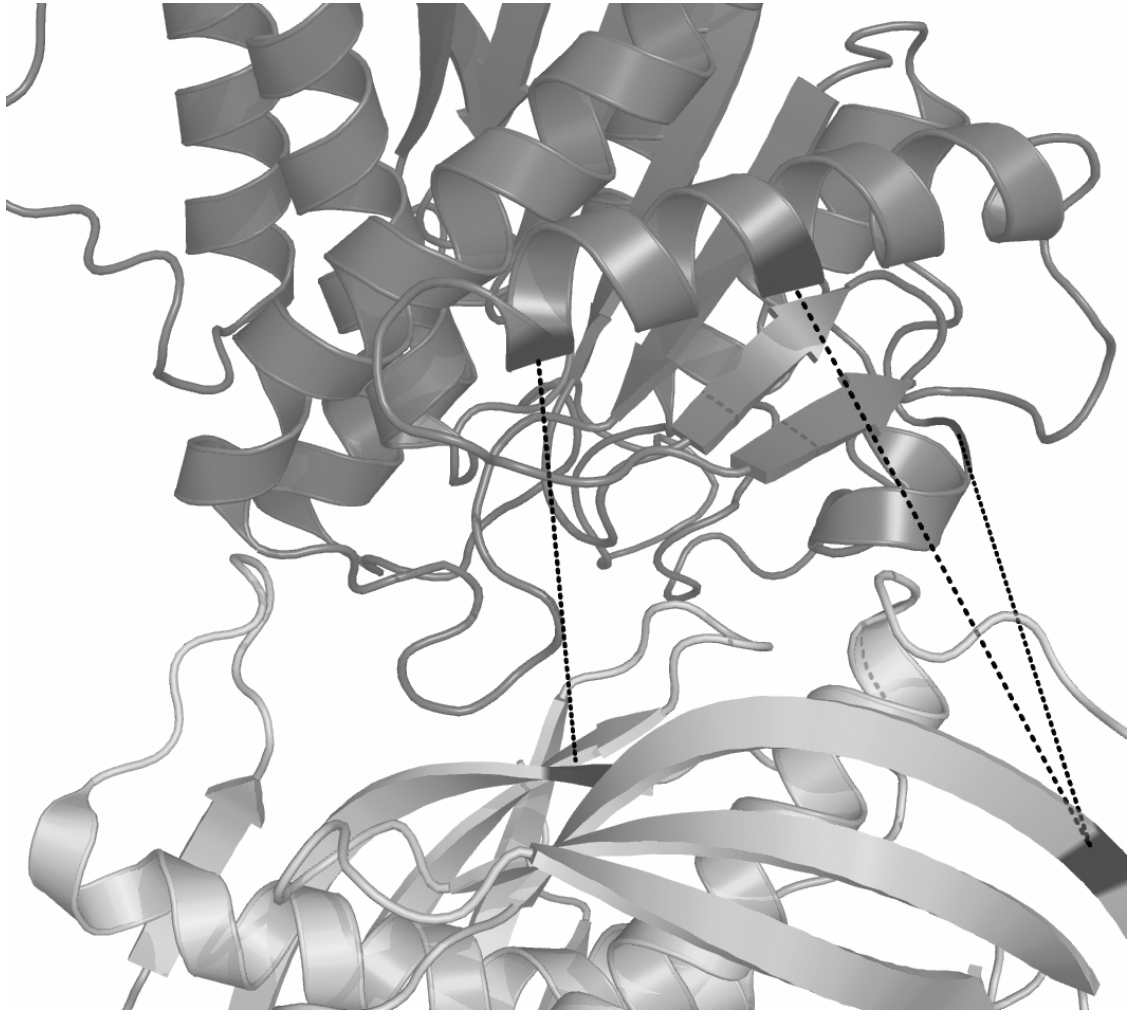
Figure 6

Figure 7

Figure 8

Figure 9

References

1. Mann, M., and Talbo, G. (1996) Developments in matrix-assisted laser desorption ionization peptide mass spectrometry *Curr. Opin. Biotechnol.* **7,** 11-19.

2. McLafferty, F. W., Fridriksson, E. K., Horn, D. M., Lewis, M. A., and Zubarev, R. A. (1999) Biochemistry - Biomolecule mass spectrometry *Science* **284,** 1289-1290.

3. Pearson, K. M., Pannell, L. K., and Fales, H. M. (2002) Intramolecular cross-linking experiments on cytochrome c and ribonuclease A using an isotope multiplet method *Rapid Commun. Mass Spectrom.* **16,** 149-159.

4. Dihazi, G. H., and Sinz, A. (2003) Mapping low-resolution three-dimensional protein structures using chemical cross-linking and Fourier transform ion-cyclotron resonance mass spectrometry *Rapid Commun. Mass Spectrom.* **17,** 2005-2014.

5. Kruppa, G. H., Schoeniger, J., and Young, M. M. (2003) A top down approach to protein structural studies using chemical cross-linking and Fourier transform mass spectrometry *Rapid Commun. Mass Spectrom.* **17,** 155-162.

6. Chen, X. H., Chen, Y. H., and Anderson, V. E. (1999) Protein cross-links: Universal isolation and characterization by isotopic derivatization and electrospray ionization mass spectrometry *Anal. Biochem.* **273,** 192-203.

7. Mouradov, D., Craven, A., Forwood, J. K., Flanagan, J. U., Garcia-Castellanos, R., Gomis-Ruth, F. X., Hume, D. A., Martin, J. L., Kobe, B., and Huber, T.

(2006) Modelling the structure of latexin-carboxypeptidase A complex based on chemical cross-linking and molecular docking *Protein Eng., Des. Sel.* **19,** 9-16.

8.   Young, M. M., Tang, N., Hempel, J. C., Oshiro, C. M., Taylor, E. W., Kuntz, I. D., Gibson, B. W., and Dollinger, G. (2000) High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry *Proc. Natl. Acad. Sci. U. S. A.* **97,** 5802-5806.

9.   Sinz, A. (2006) Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions *Mass Spectrom. Rev.* **25,** 663-682.

10. Trester-Zedlitz, M., Kamada, K., Burley, S. K., Fenyo, 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.

11. Sinz, A., Kalkhof, S., and Ihling, C. (2005) Mapping protein interfaces by a trifunctional cross-linker combined with MALDI-TOF and ESI-FTICR mass spectrometry *J. Am. Soc. Mass Spectrom.* **16,** 1921-1931.

12. Duan, X., and Sheardown, H. (2006) Dendrimer crosslinked collagen as a corneal tissue engineering scaffold: Mechanical properties and corneal epithelial cell interactions *Biomaterials* **27,** 4608-4617.

13. Bragg, P. D., and Hou, C. (1975) Subunit Composition, Function, and Spatial Arrangement in Ca2+-Activated and Mg2+-Activated Adenosine Triphosphatases of Escherichia-Coli and Salmonella-Typhimurium *Arch. Biochem. Biophys.* **167,** 311-321.

14. Lomant, A. J., and Fairbanks, G. (1976) Chemical Probes of Extended Biological Structures - Synthesis and Properties of Cleavable Protein Cross-Linking Reagent [Dithiobis(Succinimidyl-S-35 Propionate) *J. Mol. Biol.* **104,** 243-261.

15. Swaim, C. L., Smith, J. B., and Smith, D. L. (2004) Unexpected products from the reaction of the synthetic cross-linker 3,3 '-dithiobis(sulfosuccinimidyl propionate), DTSSP with peptides *J. Am. Soc. Mass Spectrom.* **15,** 736-749.

16. Hunter, M. J., and Ludwig, M. L. (1962) Reaction of Imidoesters with Proteins and Related Small Molecules *J. Am. Chem. Soc.* **84,** 3491-&.

17. Liu, S. C., Fairbanks, G., and Palek, J. (1977) Spontaneous, Reversible Protein Cross-Linking in Human Erythrocyte-Membrane - Temperature and Ph-Dependence *Biochemistry* **16,** 4066-4074.

18. Lutter, L. C., Bode, U., and Kurland, C. G. (1974) Ribosomal-Protein Neighborhoods .3. Cooperativity of Assembly *Mol. Gen. Genet.* **129,** 167-176.

19. Giron-Monzon, L., Manelyte, L., Ahrends, R., Kirsch, D., Spengler, B., and Friedhoff, P. (2004) Mapping protein-protein interactions between MutL and MutH by cross-linking *J. Biol. Chem.* **279,** 49338-49345

20. Partis, M. D., Griffiths, D. G., Roberts, G. C., and Beechey, R. B. (1983) Cross-Linking of Protein by Omega-Maleimido Alkanoyl N-Hydroxysuccinimido Esters *J. Protein Chem.* **2,** 263-277.

21. Schilling, B., Row, R. H., Gibson, B. W., Guo, X., and Young, M. M. (2003) MS2Assign, automated assignment and nomenclature of tandem mass spectra of chemically crosslinked peptides *J. Am. Soc. Mass Spectrom.* **14,** 834-850.

22. Muller, D. R., Schindler, P., Towbin, H., Wirth, U., Voshol, H., Hoving, S., and Steinmetz, M. O. (2001) Isotope tagged cross linking reagents. A new tool in mass spectrometric protein interaction analysis *Anal. Chem.* **73,** 1927-1934.

23. Davidson, W. S., and Hilliard, G. M. (2003) The spatial organization of apolipoprotein A-I on the edge of discoidal high density lipoprotein particles - A mass spectrometry study *J. Biol. Chem.* **278,** 27199-27207.

24. Bennett, K. L., Kussmann, M., Bjork, P., Godzwon, M., Mikkelsen, M., Sorensen, P., and Roepstorff, P. (2000) Chemical cross-linking with thiol-cleavable reagents combined with differential mass spectrometric peptide mapping - A novel approach to assess intermolecular protein contacts *Protein Sci.* **9,** 1503-1518.

25. Sinz, A., and Wang, K. (2001) Mapping protein interfaces with a fluorogenic cross-linker and mass spectrometry: Application to nebulin-calmodulin complexes *Biochemistry* **40,** 7903-7913.

26. Itoh, Y., Cai, K., and Khorana, H. G. (2001) Mapping of contact sites in complex formation between light-activated rhodopsin and transducin by covalent crosslinking: Use of a chemically preactivated reagent *Proc. Natl. Acad. Sci. U. S. A.* **98,** 4883-4887.

27. Krause, E., Wenschuh, H., and Jungblut, P. R. (1999) The dominance of arginine-containing peptides in MALDI-derived tryptic mass fingerprints of proteins *Anal. Chem.* **71,** 4160-4165.

28. Beardsley, R. L., and Reilly, J. P. (2002) Optimization of guanidination procedures for MALDI mass mapping *Anal. Chem.* **74,** 1884-1890.

29. Schulz, D. M., Ihling, C., Clore, G. M., and Sinz, A. (2004) Mapping the topology and determination of a low-resolution three-dimensional structure of the calmodulin-melittin complex by chemical cross-linking and high-resolution FTICRMS: Direct demonstration of multiple binding modes *Biochemistry* **43,** 4703-4715.

30. Pallares, L., Bonet, R., Garcia-Castellanos, R., Ventura, S., Aviles, F. X., Vendrell, J., and Gomis-Ruth, F. X. (2005) Structure of human carboxypeptidase A4 with its endogenous protein inhibitor, latexin *Proc. Natl. Acad. Sci. U. S. A.* **102,** 3978-3983.