Accepted Article

Title: A Click-chemistry based enrichable cross-linker for structural and protein interaction analysis by mass spectrometry

Authors: Michael Stadlmeier, Leander Simon Runtsch, Filipp Streshnev, Martin Wühr, and Thomas Carell

WILEY-VCH

COMMUNICATION

A Click-chemistry based enrichable cross-linker for structural and protein interaction analysis by mass spectrometry

Michael Stadlmeier[a],+, Leander Simon Runtsch[a],+, Filipp Streshnev[a], Martin Wühr[b] and Thomas Carell[a],*

Abstract: *Mass spectrometry (MS) is the method of choice for the characterization of proteomes. Most proteins operate in protein complexes, where their close association modulates their function. However, with standard MS analysis the information of protein-protein interactions is lost and no structural information is retained. In order to gain structural and interactome data, new cross-linking reagents are needed that freeze inter- and intramolecular interactions. Here, we report the development of a new reagent, which has several features that enable highly sensitive cross-linking MS. The reagent enables enrichment of cross-linked peptides from the majority of background peptides to facilitate efficient detection of low abundant cross-linked peptides. Due to the special cleavable properties, the reagent can be utilised for MS²- and potentially for MS³-experiments. Thus our new cross-linking reagent, in combination with high end MS, should enable sensitive analysis of interactomes, which will help researchers to obtain important insights into cellular states in health and diseases.*

Proteins need to interact with other proteins in order to form functional complexes. In many cases, protein function inside cells can not be understood without the information about the protein structure and the knowledge in which complex the protein is situated.[1-2] This is for example of paramount importance for proteins that modulate epigenetic information on DNA. Almost all of these chromatin modifying proteins require intensive interaction with metabolic enzymes, which provide the cofactors needed for histone acetylation, deacetylation or methylation and demethylation.[3-5] This showcases the need to study the complex environment of a given protein in order to analyse its function and activity state. Protein cross-linking in combination with analysis by mass spectrometry (XL-MS) is ideally suited as a method to gain information about protein structure as well as the composition of protein complexes.[6-9] For XL-MS, specialised chemical reagents, so called cross-linkers, are required, which are able to covalently connect protein residues that are in close proximity, e. g. interacting in a complex.[10-11] Characterization of the cross-linked peptides by mass spectrometry, however, poses a formidable challenge for two reasons. First, cross-link-identification has to be achieved by analyzing fragment ions of not only one, but two connected peptides, thereby massively complicating MS²-spectra.

[a] Dr. M. Stadlmeier, M. Sc. L. S. Runtsch, M. Sc. F. Streshnev, Prof. Dr. T. Carell
Department of Chemistry, Ludwig-Maximilians-Universität München, Butenandtstr. 5-13, 81377 Munich, Germany

[b] Prof. Dr. Martin Wühr
Department of Molecular Biology & the Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08544 (USA)

[*] corresponding author Thomas.Carell@lmu.de

[+] These authors contributed equally to this work.
Supporting information for this article is given via a link at the end of the document.

Figure 1. The novel cross-linker cliXlink (1). A) Structure of the cliXlink-reagent (**1**). B) Workflow for Cross-linking-MS-Experiments. After addition of **1**, protein sites in close proximity are covalently linked. The generated cross-links can be functionalised with a copper-catalysed Huisgen reaction (CuAAC). Excess reagents are removed by acetone-precipitations. After enzymatic digestion and enrichment *via* magnetic streptavidin-beads, cross-linked peptides were cleaved off under mild conditions and subsequently analysed by LC-MS². C) Fragmentation pathways a) and b) of the cross-linked peptides, forming two peptide pairs with a characteristic Δm_{rep} of 31.9721 u.

Second, the cross-linked peptide species have only a very low abundance and are part of a peptide mixture that is overwhelmingly dominated by non-cross-linked peptides. In many cases this leads to a dramatic information loss that hampers accurate cross-link

COMMUNICATION

identification and therefore interactome analysis. A few cross-linking reagents were developed, which introduce MS-cleavable groups and the possibility for XL-enrichment to tackle these problems.[12-18] Here, we report the development of a new cross-linking reagent that is enrichable and which has cleavage properties that allow accurate cross-link identification. The designed cliXlink (**1**) is depicted in Fig. 1. The cell permeable reagent (Fig 1A, Supp. Fig. 1) features two succinimidylester units that can react with nucleophiles in proteins and are spaced roughly 9 Å apart, thus enabling the fixation of short distances to gain structural information of proteins and to capture proteins in close proximity to each other. Efficient MS-cleavability is ensured by the well-established sulfoxide group, which can be cleaved with low-energy collision-induced dissociation (CID) conditions prior to the peptide fragmentation. In addition, an alkyne unit allows to attach an enrichment moiety, e.g. biotin, to the cross-linked sites with the help of a Cu(I) catalysed Huisgen reaction.[19]

Application of the crosslinker can follow a workflow as depicted in Fig. 1B. It involves addition of reagent **1** to a complex proteome, fixating the native state distance information of peptides in close proximity and preserving them throughout the further workflow for MS-analysis. Attachment of the affinity group for enrichment is carried out after cross-linking to avoid interference of bulky enrichment groups. By modification of the cross-linked peptides on the protein level, excess small molecule reagents can readily be removed by acetone precipitation. This is followed by enzymatic digestion of the proteins. The cross-linked peptides are in this way labelled, e.g. with biotin, which allows for their enrichment. Afterwards, they are analysed by mass spectrometry. Because the individual peptide masses in a cross-link are not immediately accessible, cross-link assignment is difficult, especially in complex samples. This requires to use MS-cleavable reagents, which facilitate MS²-identification by forming specific fragment ions.[20-21] At best, the reagent should also enable MS³-experiments, which requires that the cross-linker cleaves before the peptides. This allows to separate the peptides for individual MS³-based identification. Our reagent **1** contains β-hydrogens on both sides of the sulfoxide, so that the fragmentation occurs in two directions, as depicted in Fig. 1C. This leads to a clean separation of the peptides (α, β) and it generates two fragments for each peptide; an alkene and a sulfenic acid fragment, with the latter forming a thial upon water loss. This provides two mass pairs with a characteristic $\Delta m/z \approx 32$. Importantly, we noticed that the fragmentation pathway a) (Fig. 1C) predominates. Therefore, for the α-peptide the alkene- and for the β-peptide the thial-fragments are the main cleavage products. Due to the asymmetric cleavage properties, most of the signal intensity is retained on one fragment per peptide, which should provide excellent sensitivity in MS³-experiments.

The synthesis of the reagent **1** is straight forward as shown in Scheme 1. Starting point is the oxidised disulfide dimer of homocystein **2**, which is reacted with 4-pentynoic acid to disulfide **3**. Reductive cleavage of the disulfide and alkylation of the generated thiols with methyl 3-bromopropionate generates sulfide compound **4**. Saponification of both methyl esters to **5** and conversion of **5** into activated bissuccinimidyl-ester **6**, followed by oxidation of the thioether to the sulfoxide furnishes cliXlink reagent **1** in a total yield of 16%. Of particular importance is that the reagent is very pure and that the reactive ester units are in place at both sides. Partial hydrolysis needs to be avoided. This was ensured by a final precipitation purification. Reagent **1** was dissolved in a mixture of ethyl acetate and dichloromethane and precipitated upon addition of hexanes.
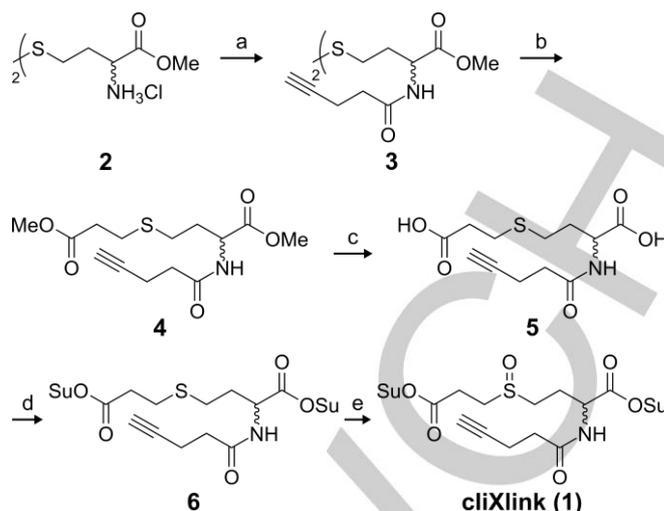
Scheme 1. Synthesis of cliXlink (1). a) 4-pentynoic acid, EDC·HCl, HOBt·H₂O, NEt₃, DMF, rt, 15 h, 71%. b) over two steps: 1) TCEP·HCl, NaHCO₃, DMF, H₂O, rt, 3 h; 2) Methyl 3-bromopropionate, 45 °C, 44 h, 55%. c) LiOH, THF, H₂O, 0 °C → rt, on, 91%. d) NHS, pyridine, trifluoroacetic anhydride, MeCN, 0 °C → rt, on, 78%. e) *m*CPBA, AcOEt, rt, 30 min, 57%.

We next investigated the mass spectrometric properties of the new reagent **1**. To this end, we added the reagent to the commonly used model protein bovine serum albumin (BSA). We followed the workflow depicted in Fig. 1B without performing the enrichment step. In short, after addition of the reagent to the protein solution and reaction for 1 h at room temperature and physiological pH, we precipitated the protein with acetone, resuspended in buffer (see SI) and digested the protein subsequently with a mixture of trypsin and Lys-C.

The obtained peptide mixture was desalted with C₁₈-tips and analysed by HPLC-MS². The resulting data are depicted in Fig. 2. Fig. 2A shows two cross-linked BSA peptides (α + β) as an example. After determining the exact mass of the intact cross-link (*m/z* 677.1, Fig. 2B), we performed both CID and HCD fragmentation on the precursor to evaluate the fragmentation properties (Fig 2C). The more selective CID fragmentation (resonance excitation of the cross-link ion, top spectrum) at a low normalised collision energy of 25% provides as expected just a small set of intense signals. The two prominent signal pairs with a Δ*m/z* of 32 (Δm_rep) are obtained due to cross-linker fragmentation which results in a thial (α/β thial) and alkene (α/β alkene) fragment for each peptide. As mentioned before, fragmentation pathway a) is preferred, leading to an asymmetric intensity distribution. The dominant formation of the expected intact, separated peptides makes the reagent consequently amenable for more sophisticated MS³ experiments.

For our study we used HCD fragmentation. This method provides simultaneous cross-linker cleavage and formation of the peptide fragments required for identification (Fig 2C, bottom spectrum). To help with cross-link/peptide recognition, it is important that the HCD fragmentation still provides the alkene and thial fragments. The HCD data therefore allow to identify the peptides already by MS².

COMMUNICATION

61 BSA unique cross-links
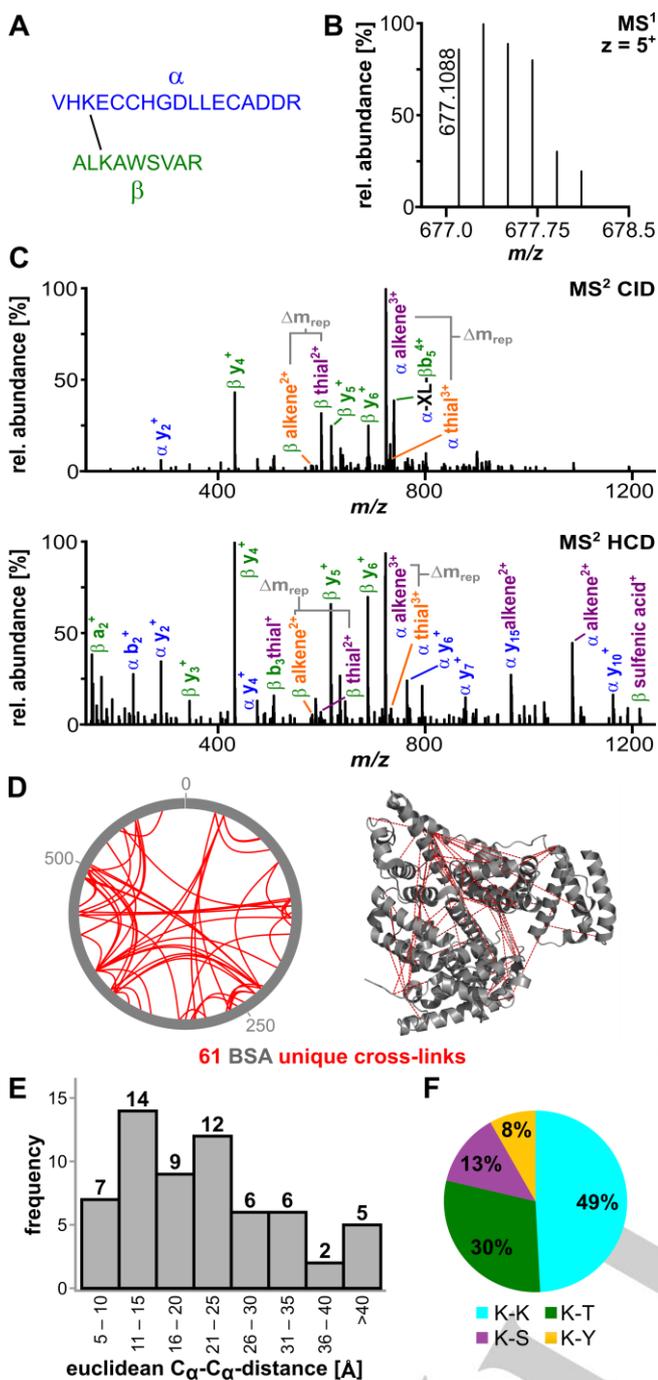
Figure 2. Analysis of cross-linked BSA prior to enrichment. A) Sequences of a representative cross-link within BSA: Peptide α: VHKECCHGDLLECADDR (IAA-modified on Cys) and β: ALKAWSVAR. B) The isotope envelope of the 5+-charged precursor is shown. C) Fragmentation of the cross-linked peptides under CID- (25%, top spectrum) and stepped HCD-conditions (25%/30%/32%, bottom spectrum) showing the product ions of the sulfoxide-cleavage. The different fragmentation pathways are shown in orange and purple (see Fig. 1). D) MS²-identification of BSA cross-link-sites prior to enrichment. The xVis-representation shows the positions of cross-linked amino acids as red lines (left). The corresponding C_α-C_α-distances are depicted in the BSA crystal structure (pdb: 4F5S[22]). E) Histogram of the measured euclidean C_α-C_α-distances in the crystal structure with the majority being below 40 Å and a mean measured distance of 22.1 Å. F) Distribution of the identified cross-linking sites between Lys-Lys, Lys-Thr, Lys-Ser and Lys-Tyr.

Using this method, we analysed the data with the freely available MeroX-software, strictly filtering all cross-link identifications (score > 50, FDR 1%) and requiring the presence of at least 3 out of 4 ions from the two characteristic mass pairs.[23-24] Without exploiting the possibility for CuAAC-based enrichment, we were able to identify 61 unique BSA cross-links in a single measurement, shown as an xVis representation (Fig. 2D).[25] This result is representative for measurements conducted with purified BSA in our lab. Additionally, we depicted the found cross-links in the BSA crystal stucture and noted that the majority of measured distances are reasonable. Importantly, the cross-links show a mean C_α-C_α distance between the cross-linked amino acids of 22.1 Å and a distance distribution which is in perfect agreement of what is expected for the cross-linking with a reagent that spaces the active esters by about 9 Å (Fig. 2E, Supp. Fig. 2)[26] and taking into account side chain lengths and protein molecular dynamics.[27] Cross-linking occurred mostly between two Lys residues (49%), but we also detected Lys-Thr (30%), Lys-Ser (13%) and Lys-Tyr (8%) connections in agreement with the reactivity of the succinimidyl esters (Fig. 2F).[28]

This success allowed us to validate the new cross-linker in a more complex environment. We particularly wanted to show the additional value of the CuAAC-based enrichment possibility. For the experiment we again cross-linked the BSA protein, precipitated with acetone, redissolved the cross-linked protein and subsequently performed the click reaction (Supp. Fig. 3A) by adding biotin-disulfide-azide **7** (Fig. 3A) and CuSO₄/THPTA followed by reduction of Cu(II) to Cu(I) upon addition of sodium ascorbate. After 1 h at room temperature, we performed another acetone precipitation to remove excess biotin azide. We then added trypsin and Lys-C for digestion.
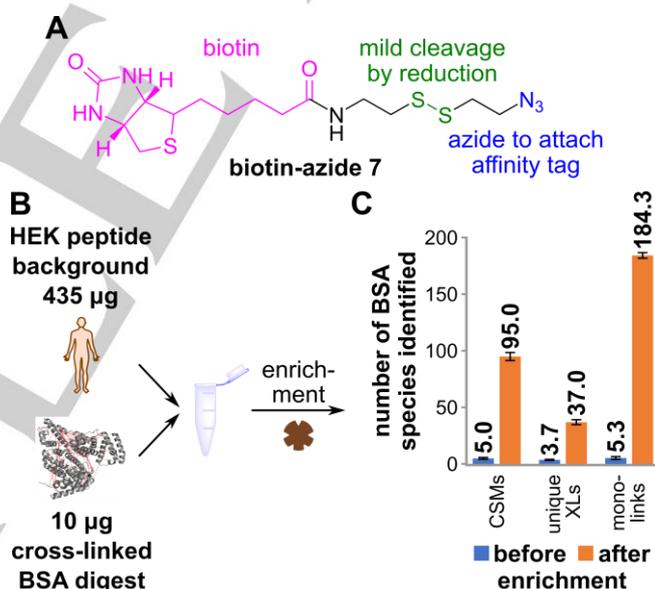
Figure 3. Enrichment of cross-linked BSA from a complex sample. A) Structure of the utilised biotin-azide **7**, with the biotin group (pink), the disulfide-bond (green) and the azide-moiety (blue). Upon reduction of the disulfide, the biotin moiety is removed from the cross-link. B) Depiction of the spike-in-experiment. 10 µg of cross-linked and CuAAC-modified BSA digest were added to 435 µg of HEK digest as a complex background. Magnetic streptavidin-beads were used to enrich the cross-linked peptides. C) Result of the HPLC-MS²-analysis showing effect of enrichment. The experiment was done in technical triplicates, numbers show the mean value and the standard deviation is indicated by the error bar.

COMMUNICATION

These peptides we next combined with a protein digest obtained from a HEK cell extract (435 µg of protein, Fig. 3B). This creates a massive background of non-cross-linked peptides. When we now analysed the cross-links within this large background (Fig. 3C), we identified only a very small number of cross-link spectral matches (mean CSMs = 5.0), unique cross-linked sites (mean unique XLs = 3.7) as well as monolinks (mean = 5.3), where one succinimidylester hydrolysed before reaction with the protein. However, when we performed the enrichment using streptavidin coated magnetic beads, the number of cross-link identifications improved dramatically. Following incubation of the peptide mixture with the magnetic particles, extensive washing (6x) of the beads with buffer (see SI) and mild, reductive cleavage of the disulfide to liberate the "captured" cross-links and thereby removing the biotin moiety (Supp. Fig. 3B), we now on average detected 95.0 CSMs, 37.0 unique XLs together with 184.3 monolinks. The number of identified unique XLs is lower than for purified BSA (Fig. 2D), which could be explained by inefficiencies in the CuAAC reaction or interference of the non-crosslinked complex background. However, this result shows that our new cross-linker **1** is able to enrich cross-linked peptides in a vast excess of unmodified peptides, facilitating the analysis of complex samples with low abundant cross-links. Originally, we were sceptical about the asymmetric structure of cliXlink (**1**). The data, however, show that despite this, a large number of cross-links is identified. In summary, our new cross-linker **1** combines the following advantages: first, the size of the reagent is ideally suited to gain valuable distance information for structural proteomics. In addition, the molecule is cell permeable and the alkyne unit does not cause major interference with the cross-linking reaction. Furthermore, the robust, efficient sulfoxide fragmentation simplifies cross-link identification. The possibility to functionalise peptides cross-linked with **1** *via* Click chemistry allows to employ diverse modifications and enrichment strategies enabling analysis of low abundant cross-links. To conclude, our reagent cliXlink (**1**) now paves the way for complex interactome analysis using MS^2 and also MS^3 experiments.

[1] I. N. Berezovsky, E. Guarnera, Z. Zheng*, et al.*, *Curr. Opin. Struc. Biol.* **2017**, *42*, 67-74.
[2] H. Currinn, T. Wassmer, *Biochem. Soc. T.* **2016**, *44*, 185-190.
[3] S. Allard, R. T. Utley, J. Savard*, et al.*, *EMBO J.* **1999**, *18*, 5108-5119.
[4] Y. Xue, J. Wong, G. T. Moreno*, et al.*, *Mol. Cell* **1998**, *2*, 851-861.
[5] M. Sauvageau, G. Sauvageau, *PLOS Biol.* **2008**, *6*, e113.
[6] A. Leitner, T. Walzthoeni, A. Kahraman*, et al.*, *Mol. Cell. Proteomics* **2010**, *9*, 1634-1649.
[7] A. Sinz, *Angew. Chem. Int. Ed.* **2018**, *57*, 6390-6396.
[8] J. Rappsilber, *J. Struct. Biol.* **2011**, *173*, 530-540.
[9] J. D. Chavez, J. E. Bruce, *Curr. Opin. Chem. Biol.* **2019**, *48*, 8-18.
[10] F. J. O'Reilly, J. Rappsilber, *Nat. Struct. Mol. Biol.* **2018**, *25*, 1000-1008.
[11] A. Leitner, M. Faini, F. Stengel*, et al.*, *Trends Biochem. Sci.* **2016**, *41*, 20-32.
[12] A. Kao, C. L. Chiu, D. Vellucci*, et al.*, *Mol. Cell. Proteomics* **2011**, *10*, M110 002212.
[13] M. Q. Muller, F. Dreiocker, C. H. Ihling*, et al.*, *Anal. Chem.* **2010**, *82*, 6958-6968.
[14] A. M. Burke, W. Kandur, E. J. Novitsky*, et al.*, *Org. Biomol. Chem.* **2015**, *13*, 5030-5037.
[15] X. Tang, J. E. Bruce, *Mol. Biosyst.* **2010**, *6*, 939-947.
[16] R. Huang, W. Zhu, Y. Wu*, et al.*, *Chem. Sci.* **2019**, *10*, 6443-6447.
[17] M. Rey, M. Dupre, I. Lopez-Neira*, et al.*, *Anal. Chem.* **2018**, *90*, 10707-10714.
[18] B. Steigenberger, R. J. Pieters, A. J. R. Heck*, et al.*, *ACS Cent. Sci.* **2019**, *5*, 1514-1522.
[19] M. Ehrlich, M. J. Gattner, B. Viverge*, et al.*, *Chem. Eur. J.* **2015**, *21*, 7701-7704.
[20] F. Liu, P. Lossl, R. Scheltema*, et al.*, *Nat. Commun.* **2017**, *8*, 15473.
[21] O. Klykov, B. Steigenberger, S. Pektas*, et al.*, *Nat. Protoc.* **2018**, *13*, 2964-2990.
[22] A. Bujacz, *Acta Crystallogr. D* **2012**, *68*, 1278-1289.
[23] M. Götze, J. Pettelkau, R. Fritzsche*, et al.*, *J. Am. Soc. Mass. Spectrom.* **2015**, *26*, 83-97.
[24] M. Götze, C. Iacobucci, C. H. Ihling*, et al.*, *Anal. Chem.* **2019**, *91*, 10236-10244.
[25] M. Grimm, T. Zimniak, A. Kahraman*, et al.*, *Nucleic Acids Res.* **2015**, *43*, W362-369.
[26] E. D. Merkley, S. Rysavy, A. Kahraman*, et al.*, *Protein Sci.* **2014**, *23*, 747-759.
[27] Z. Orban-Nemeth, R. Beveridge, D. M. Hollenstein*, et al.*, *Nat. Protoc.* **2018**, *13*, 478-494.
[28] S. Kalkhof, A. Sinz, *Anal. Bioanal. Chem.* **2008**, *392*, 305-312.
[29] Y. Perez-Riverol, A. Csordas, J. Bai*, et al.*, *Nucleic Acids Res.* **2019**, *47*, D442-450.