**Supplementary Material**

# Chemical cross-linking and mass spectrometry as a *low-resolution* protein *structure determination technique*

Pragya Singh, Alexandre Panchaud, David R. Goodlett

Department of Medicinal Chemistry, University of Washington, Seattle, WA

## Supplementary text

### Nomenclature

A lack of consistent nomenclature describing CXMS is apparent in the existing literature. For many chemically cross-linked entities there are no standardized names in the literature and as a consequence each group has used their own conventions. This makes assimilating information tedious for the novice and also makes it difficult for authors to report and compare their findings. The only effort made towards a consistent nomenclature is reported by Schilling *et al.*, who introduced specific nomenclature for cross-linked peptides[1]. However, this nomenclature scheme has not been adopted consistently in the literature. Here we use the nomenclature suggested by Schilling *et al.* and take it a step further by appending it with nomenclature for several other chemical entities that are referenced repeatedly in the literature. This proposed nomenclature is for the most part based on the terms already reported, but not consistently used, in the literature. We believe that use of a consistent nomenclature will benefit both authors and readers in the scientific community and suggest adoption of this modified "Schilling" nomenclature convention.

**CXMS:** The process of chemical cross-linking two amino acid residues close in space with a chemical reagent followed by mass spectrometric analysis of the whole cross-linked complex or the digested peptides. Use of CXMS can replace phrases like *chemical cross-linking coupled to mass spectrometry*, *chemical cross-linking and mass spectrometry*, or *chemical cross-linking followed by mass spectrometry*

**Type 0 cross-links:** Peptides modified with the hydrolyzed cross-linker. These have been referred to as *dead-end*, *monolinks*, or *monoderivatization product*.

**Type 1 cross-links:** A loop-like structure created when two arms of a chemical cross-linker react with two amino acid residues close in primary sequence of a peptide. These are often referred to as *intra-peptide cross-links*, *loop-links*, or *cyclic cross-links*.

**Type 2 cross-links:** A structure where two independent peptides are chemically connected with a chemical cross-linker creating an "H"-shaped structure. Type 2 cross-links can be intra-protein (within the same protein) or inter-protein (between two proteins). *Conjugates* or *cross-linked peptides* are examples of names in use for Type 2 cross-links.

**Type 2 α-peptide:** The longer peptide chain of a Type 2 cross-linked peptide.

**Type 2 β-peptide:** The shorter peptide chain of a Type 2 cross-linked peptide.

**Non-cross-linked fragment ions:** Fragment ions in the tandem mass spectrum of type 2 cross-link that do not contain cross-linked amino acids. For example b-ions located on the N-terminus of the cross-linked site or y-ions that are located on the C-terminus of the cross-linked site. These fragment ions are denoted as b- and y-ions the same as for linear peptide fragment ions. Non-crosslinked fragment ions are also referred to as *common ions*.

**Cross-linked fragment ions:** Fragment ions in the tandem mass spectrum of type 2 cross-linked peptides that contain cross-linked amino acids. We propose to represent these ions by the

conventional b and y nomenclature followed by the superscript x (such as $b^x$ and $y^x$) to indicate that they carry an extra mass of the cross-linker and the second peptide in addition to their usual fragment ion mass.

**Cross-linkers - from homo- to hetero-functional**

***Chemical specificity.*** Long before mass-spectrometric analysis became routine for identification of proteins, chemical cross-linking methods were being used to study protein interactions. As early as the 1940s, experimental evidence was presented to illustrate that formaldehyde could cross-link amino-acid side chains in proteins[2]. It was the first chemical compound used to cross-link proteins and remains one of the most popular cross-linking reagents for *in vivo* studies[3, 4]. Formaldehyde can cross-link free amino groups as well as the side chains of Histidine, Cysteine, Tyrosine, Tryptophan, and Phenylalanine[5] making it non-specific. However, most other cross-linking reagents are directed towards specific amino acid side chains (**Supplementary Figure 1A**). A majority of cross-linking reagents reported in the literature and those available commercially have functional groups targeting one of the following chemical groups in proteins: amine, carboxyl, and sulfhydryl groups[6, 7]. Cross-linking reagents reactive towards primary amino groups are by far the most common (**Supplementary Figure 1B**). These reagents most often contain an *N*-hydroxysuccinimide (NHS) ester or an imidoester group to acylate primary amino groups[8, 9]. At physiological pH NHS-esters have better cross-linking efficiency compared to imidoesters[10]. Although NHS-esters were previously believed to react only with primary amines of lysines and N-termini of proteins, it has been shown that they are also reactive towards serine and tyrosine residues as well as contaminant ammonium ions in buffer solutions[11]. Cross-linking reagents can also contain a carbodiimide functional group that renders carboxylic group reactive towards primary amino group to form an amide bond[12]. Carbodiimides are known as "zero-length" cross-linkers because they form a direct chemical bond between the two amino acids leaving no chemical trace of the reagent (**Supplementary Figure 1C**). Sulfhydryl groups are targeted by cross-linkers containing maleimide, haloacetyl

and pyridyl disulfide reactive groups (**Supplementary Figure 1D**)[13-15]. Typically these reagents have limited utility because they can disrupt native disulfide bonds in proteins, distorting their native three-dimensional structure. Carbonyl groups can also be targeted by cross-linking reagents. Such reagents are especially useful in forming conjugates between antibodies and proteins by reaction with the carbohydrate moieties of the antibody. Because carbonyls do not readily exist in proteins, mild oxidation of sugar glycols is first performed using sodium *meta*-periodate for the conversion to aldehydes or ketones. Subsequent reaction is then possible with hydrazide compound to form a hydrazone bond (**Supplementary Figure 1E**). While these specific reagents have the advantage of targeting defined residues in proteins, photo-reactive cross-linking reagents, which are fairly non-specific, can be made to react at selected activation times. Photo-reactive reagents insert themselves into neighboring residues in response to irradiation by UV light, allowing conjugation reaction to be easily controlled in time (**Supplementary Figure 1F**). Detailed description of the cross-linkers mentioned in this section can be found in the book by S. S. Wong[7] and several review articles[10, 16, 17]. The technical handbook edited by Pierce on cross-linking reagents is also a valuable source of information[6].

***Number of reactive groups.***

i) Bifunctional reagents: In order to study protein interactions using chemical cross-linking, the cross-linkers need to have at least two chemically reactive moieties. These reactive centers can either be identical, making them homobifunctional, or they can be reactive towards two different functional groups making them heterobifunctional (**Supplementary Figure 1A**). For heterobifunctional cross-linking reagents, having specific side chain reactivity such as amino or sulfhydryl functional groups on one end and a photoreactive group on the other is a desirable configuration. Homobifuntional

cross-linking reagents can be used in a convenient single-step reaction protocol. However, this can give rise to unwanted, non-specific conjugations, since all reagents are added to the reaction mixture simultaneously. Alternatively, heterobifunctional reagents can go through two-step conjugation, producing more specific cross-links. In the two-step protocol the more labile functional groups in the heterobifunctional reagent can react first with a protein to form an 'activated' intermediate. After removal of the excess un-reacted reagent, this activated intermediate can be reacted with the second protein to form the cross-linked complex. When characterizing higher-order oligomers, however, both homo- and hetero-bifunctional cross-linkers have to be used in a one-step procedure.

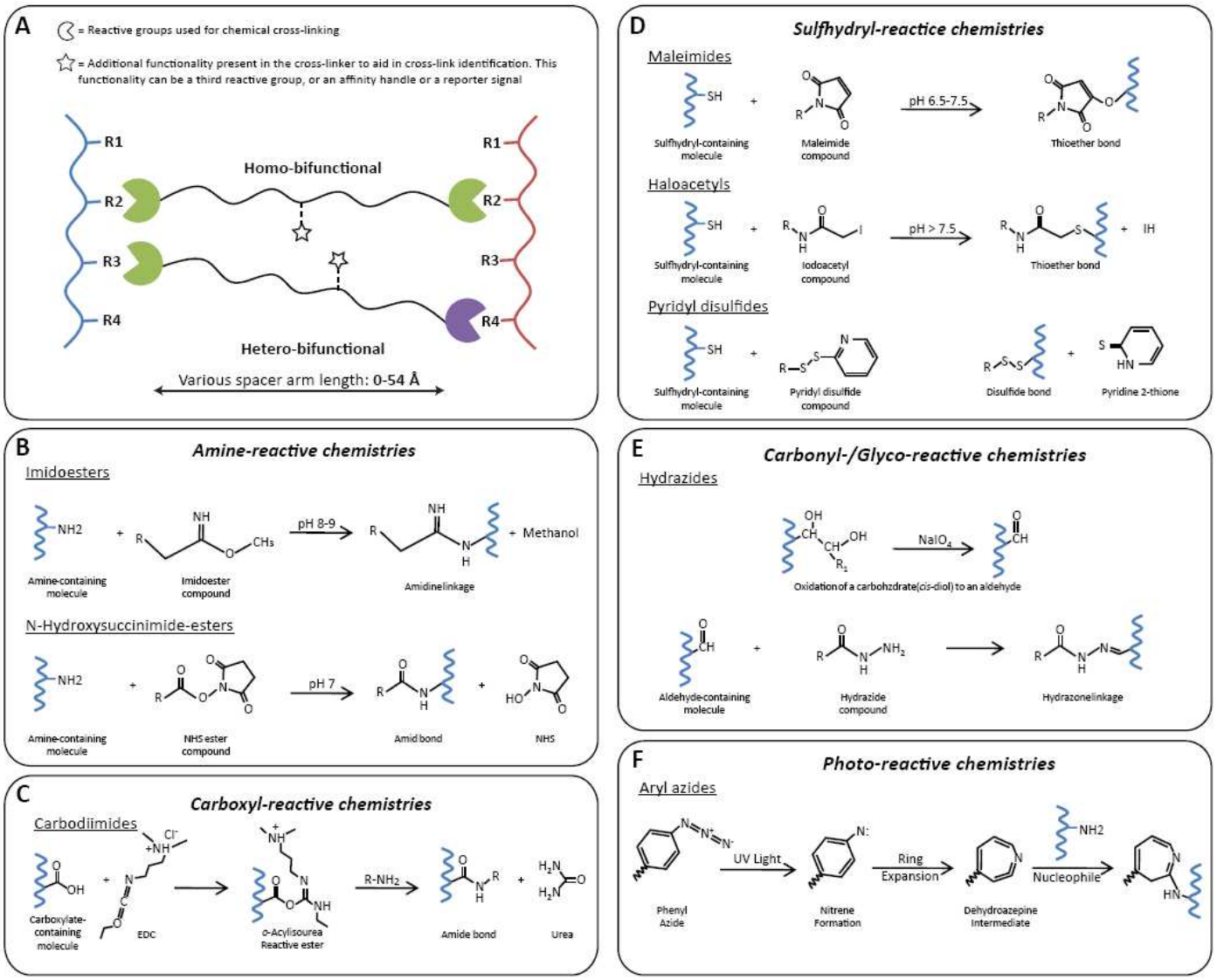
ii) Trifunctional reagents: Cross-linking reagents consisting of three reactive arms are known as trifunctional cross-linkers. For these reagents, the third arm can either have an additional functional group to link to a third peptide (e.g. Tris[2-maleimidoethyl]amine (TMEA), Tris-succinimidyl aminotriacetate (TSAT); Pierce, Rockford, IL) or it can have an affinity handle that can be used to purify cross-linked complexes after the cross-linking reaction is complete (Sulfo-N-hydroxysuccinimidyl-2-(6-[biotinamido]-2-(p-azido benzamido)-hexanoamido) ethyl-1,3'-dithioproprionate (sulfo-SBED); Pierce, Rockford, IL)[6].

***Spacer arm length.*** Distance spanned by the cross-linking reagent is an important criterion to consider during the selection process. Cross-linkers with short spacer arms span short-distances and target close interactions, which capture amino acids in a relatively small radius. As a consequence, the yield of cross-linked products from zero-length cross-linking reagents is typically low. These reagents spanning short distances also lead to more intra-protein cross-linking compared to the reagents with longer spacer arm lengths[18]. Thus it is possible to

vary the ratio of intra- to inter-molecular cross-links by adjusting the distance spanned by the cross-linking reagent, as well as its concentration. It is also very important to maximize the yield of cross-linked products without distorting the native structure of the protein or the complex, while keeping the formation of unwanted products to a minimum. This can only be achieved by testing a series of cross-linking reagents. Additionally, reagents with the same reactivity and similar spacer arm lengths can be used to cross-validate the initial cross-links identified. Alternatively, cross-linking studies with systematic variation in reagent spacer arm length can provide useful insights into the three-dimensional structure of proteins or protein complexes[19, 20].

Choice of cross-linking reagent is highly dependent on the system being studied and the reactive groups that are being targeted. Obviously amino acid composition and sequence, if known, are important. More important though are projections about overall hydropathy of the targeted surface (e.g. whether it is a hydrophobic cleft or lined by polar residues), where available, as this will aid selection of the best cross-linker. This information can be discerned in part by protein surface mapping using hydrogen exchange[21] and covalent labeling[22, 23]. In addition to the spacer arm length of the cross-linking reagent, other factors to consider for cross-linker selection are its water solubility, membrane permeability, and its compatibility with the reaction conditions in which the experiments are to be performed.

## Supplementary Figures:




**Supplementary Figure S1.** Reaction schemes for cross-linking reagents A) General scheme B) Amine-reactive cross-linkers C) Carboxyl-reactive cross-linkers D) Sulfhydryl-reactive cross-linkers E) Carbonyl-/Glyco-reactive cross-linkers F) Photo-reactive cross-linkers

## Supplementary Tables

**Supplementary Table S1.** Software used for analyzing CXMS data

| | | |
|---|---|---|
| **A. Software utilizing precursor ion data only** | | |
| Links (ASAP) | https://ms3d.ca.sandia.gov:11443/cms3d/portal | [24] |
| GPMAW | http://gpmaw.com/GPMAW/gpmaw.html | [25, 26] |
| MS-Bridge | http://prospector.ucsf.edu/ | [27] |
| CLPM | http://bioinformatics.ualr.edu/mbc/services/CLPM.html | [28] |
| VirtualMSLab | Available on request through Luitzen de Jong, ldk@science.uva.nl | [29] |
| FindLink | Unavailable | [30] |
| NIH-XL | Unavailable | [31] |
| MASA | Unavailable | [32] |
| **B. Software utilizing precursor as well as fragment ion data** | | |
| MS2Links (MS2Assign) | https://ms3d.ca.sandia.gov:11443/cms3d/portal | [1] |
| Pro-CrossLink | http://depts.washington.edu/ventures/UW_Technology/Express_Licenses/ProCrossLink.php | [33] |
| XLink | http://tools.proteomecenter.org/XLink.php | [18] |
| SearchXLink | Unavailable | [34, 35] |
| X-links | Available on request through James E. Bruce jimbruce@u.washington.edu | [36] |
| **C. Software utilizing data-dependent LC-MS/MS data** | | |
| X!Link | Available on request through Young Jin Lee yojlee@ucdavis.edu or yjlee@iastate.edu | [37] |
| XDB | Unavailable | [38] |
| xQuest | http://www.xQuest.org | [39] |
| Open-mod pipeline | http://www.expasy.ch/tools/popitam/ | [40] |

## References


(1) Schilling, B.; Row, R. H.; Gibson, B. W.; Guo, X.; Young, M. M. *J. Am. Soc. Mass Spectrom.* 2003, *14*, 834-850.
(2) Fraenkel-Conrat, H.; Mecham, D. K. *J. Biol. Chem.* 1949, *177*, 477-486.
(3) Orlando, V.; Strutt, H.; Paro, R. *Methods* 1997, *11*, 205-214.
(4) Vasilescu, J.; Guo, X.; Kast, J. *Proteomics* 2004, *4*, 3845-3854.
(5) Toews, J.; Rogalski, J. C.; Clark, T. J.; Kast, J. *Anal. Chim. Acta* 2008, *618*, 168-183.
(6) *Pierce: Rockford, IL,* 2003.
(7) Wong, S. S. *CRC Press: Boca Raton, FL* 1991.
(8) Hartman, F. C.; Wold, F. *J. Am. Chem. Soc.* 1966, *88*, 3890-3891.
(9) Bragg, P. D.; Hou, C. *Arch. Biochem. Biophys.* 1975, *167*, 311-321.
(10) Sinz, A. *Mass Spectrom. Rev.* 2006, *25*, 663-682.
(11) Swaim, C. L.; Smith, J. B.; Smith, D. L. *J. Am. Soc. Mass Spectrom.* 2004, *15*, 736-749.
(12) Hoare, D. G.; Koshland, D. E. *J. Am. Chem. Soc.* 1966, *88*, 2057-2058.
(13) Rector, E. S.; Schwenk, R. J.; Tse, K. S.; Sehon, A. H. *J. Immunol. Methods* 1978, *24*, 321-336.
(14) Smyth, D. G.; Blumenfeld, O. O.; Konigsberg, W. *Biochem. J.* 1964, *91*, 589–595.
(15) Carlsson, J.; Drevin, H.; Axén, R. *Biochem. J.* 1978, *173*, 723-737.
(16) Back, J. W.; de Jong, L.; Muijsers, A. O.; de Koster, C. G. *J. Mol. Biol.* 2003, *331*, 303-313.
(17) Sinz, A. *J. Mass Spectrom.* 2003, *38*, 1225-1237.
(18) Seebacher, J.; Mallick, P.; Zhang, N.; Eddes, J. S.; Aebersold, R.; Gelb, M. H. *J. Proteome Res.* 2006, *5*, 2270-2282.
(19) Novak P; Young MM; Schoeniger JS; GH., K.*: Eur J Mass Spectrom* 2003, *9*, 623-631.
(20) Jacobsen, R. B.; Sale, K. L.; Ayson, M. J.; Novak, P.; Hong, J.; Lane, P.; Wood, N. L.; Kruppa, G. H.; Young, M. M.; Schoeniger, J. S. *Protein Sci.* 2006, *15*, 1303-1317.
(21) Wales, T. E.; Engen, J. R. *Mass Spectrom. Rev.* 2006, *25*, 158-170.
(22) Konermann, L.; Tong, X.; Pan, Y. *J. Mass Spectrom.* 2008, *43*, 1021-1036.
(23) Sharp, J. S.; Becker, J. M.; Hettich, R. L. *Anal. Biochem.* 2003, *313*, 216-225.
(24) Young, M. M.; Tang, N.; Hempel, J. C.; Oshiro, C. M.; Taylor, E. W.; Kuntz, I. D.; Gibson, B. W.; Dollinger, G. *Proc. Nat. Acad. Sci.* 2000, *97*, 5802-5806.
(25) Hojrup, P. *In: Ion Formation from Organic Solvents (A. Hedin et al.), John Wiley & Sons* 1990, 61-66.
(26) Peri, S.; Steen, H.; Pandey, A. *Trends Biochem. Sci.* 2001, *26*, 687-689.
(27) Chu, F.; Shan, S.; Moustakas, D. T.; Alber, F.; Egea, P. F.; Stroud, R. M.; Walter, P.; Burlingame, A. L. *Proc. Nat. Acad. Sci.* 2004, *101*, 16454-16459.
(28) Tang, Y.; Chen, Y.; Lichti, C.; Hall, R.; Raney, K.; Jennings, S. *BMC Bioinformatics* 2005, *6*, S9.
(29) Koning, L. J.; Kasper, P. T.; Back, J. W.; Nessen, M. A.; Vanrobaeys, F.; Beeumen, J.; Gherardi, E.; Koster, C. G.; Jong, L. *FEBS J.* 2006, *273*, 281-291.
(30) Back, J. W.; Notenboom, V.; de Koning, L. J.; Muijsers, A. O.; Sixma, T. K.; de Koster, C. G.; de Jong, L. *Anal. Chem.* 2002, *74*, 4417-4422.
(31) Sinz, A.; Wang, K. *Biochemistry* 2001, *40*, 7903-7913.
(32) Wallon, G.; Rappsilber, J.; Mann, M.; Serrano, L. *EMBO J.* 2000, *19*, 213-222.
(33) Gao, Q.; Xue, S.; Doneanu, C. E.; Shaffer, S. A.; Goodlett, D. R.; Nelson, S. D. *Anal. Chem.* 2006, *78*, 2145-2149.
(34) Schnaible, V.; Wefing, S.; Resemann, A.; Suckau, D.; Bücker, A.; Wolf-Kümmeth, S.; Hoffmann, D. *Anal. Chem.* 2002, *74*, 4980-4988.
(35) Wefing, S.; Schnaible, V.; Hoffmann, D. *Anal. Chem.* 2006, *78*, 1235-1241.
(36) Anderson, G. A.; Tolic, N.; Tang, X.; Zheng, C.; Bruce, J. E. *J. Proteome Res.* 2007, *6*, 3412-3421.
(37) Lee, Y. J.; Lackner, L. L.; Nunnari, J. M.; Phinney, B. S. *J. Proteome Res.* 2007, *6*, 3908-3917.
(38) Maiolica, A.; Cittaro, D.; Borsotti, D.; Sennels, L.; Ciferri, C.; Tarricone, C.; Musacchio, A.; Rappsilber, J. *Mol. Cell. Proteomics* 2007, *6*, 2200-2211.

**(39)** **Rinner, O.; Seebacher, J.; Walzthoeni, T.; Mueller, L.; Beck, M.; Schmidt, A.; Mueller, M.; Aebersold, R. *Nat. Meth.* 2008, *5*, 315-318.**

**(40)** **Singh, P.; Shaffer, S. A.; Scherl, A.; Holman, C.; Pfuetzner, R. A.; Larson Freeman, T. J.; Miller, S. I.; Hernandez, P.; Appel, R. D.; Goodlett, D. R. *Anal. Chem.* 2008, *80*, 8799-8806.**