XLMOD: Cross-linking and chromatography derivatization reagents ontology

Authors: Gerhard Mayer

1 Ruhr University Bochum, Faculty of Medicine, Medizinisches Proteom-Center, D-44801 Bochum, Germany
**Abbreviations:**

| Abbreviation | Full Form |
|--------------|-----------|
| CID          | Collision Induced Dissociation |
| CV           | Controlled Vocabulary |
| CX-MS        | Cross-Linking-Mass Spectrometry |
| Da           | Dalton (atomic mass unit) |
| ESI          | ElectroSpray Ionization |
| HUPO-PSI     | HUman Proteome Organization – Proteomics Standards Initiative |
| IUPAC        | International Union of Pure and Applied Chemistry |
| InChI        | International Chemical Identifier |
| MS           | Mass Spectrometry |
| NMR          | Nuclear Magnetic Resonance |
| SMILES       | Simplified Molecular Input Line Entry Specification |
| XRC          | X-Ray Crystallography |
Summary/Abstract:
Mass spectrometry has experienced a rapid development since its first application for protein analysis in the 1980s. While the most common use of mass spectrometry for protein analysis is identification and quantification workflows on peptides (digested from their parent protein), there is also a rapidly growing use of mass spectrometry for “structural proteomics”. One example is the analysis of cross-linked proteins that can give valuable structural information, complementing the information gained by classical protein structure determination methods, useful for integrated methods of structure determination and modeling. For a broad and reproducible application of cross-linking mass spectrometry a standardized representation of cross-linking experimental results is necessary. This paper describes the developing and release of the xImod ontology from the HUPO-PSI. xImod contains terms for the description of reagents used in cross-linking experiments and of cross-linker related chemical modifications together with their main properties relevant for planning and performing cross-linking experiments. We also describe how xImod is used within the new release of HUPO-PSI’s mzIdentML data standard, for reporting the used cross-linking reagents and results in a consistent manner. In addition xImod contains terms for GC-MS and LC-MS derivatization reagents for specifying them in the upcoming mzTab-M and mzTab-L formats.

Keywords: Cross-linking reagents, cross-linker related chemical modifications, distance constraint, derivatization, xImod ontology, mzIdentML 1.2, mzTab

Download link: https://raw.githubusercontent.com/HUPO-PSI/mzIdentML/master/cv/XLMOD.obo
1. Introduction

Traditionally proteins investigations are separated into two domains: At first, there are plenty of experimental methods for isolation, purification, identification and quantification of proteins (e.g. Difference Gel Electrophoresis (DIGE), SDS-PAGE (Sodium Dodecyl Sulfate – Polyacrylamide Gel Electrophoresis), mass spectrometry (MS)). These aim at characterizing the proteins by their properties like e.g. their mass and sequence, their modifications, their physicochemical properties, interaction partners and so on. Second, there is the area of protein structural biology, aiming at the determination of tertiary and quaternary three-dimensional structure of the proteins, which is essential for their functional and dynamical behavior.

With the technological progress of protein mass spectrometry (MS) in the last 3 decades MS is becoming more and more a useful tool for structural biology by giving information like dynamical, spatial and accessibility constraints, which can either be combined with other existing experimental protein structure determination or with computer modelling methods [1, 2]. Especially cross-linking mass spectrometry (CX-MS) [3, 4] is able to complement traditional methods like X-ray crystallography (XRC), nuclear magnetic resonance (NMR) and cryo-electron microscopy (cryo-EM), since each of these methods has some limitations: X-ray crystallography and NMR are both time-consuming and costly methods needing a relatively large amount of pure proteins. Furthermore they are either difficult to apply for membrane proteins and analyze the protein not in their native conformation (XRC), can be used only for smaller soluble proteins (NMR) or give only medium resolution information (cryo-EM), whereas CX-MS is a relatively fast high-throughput (HT) method by which also small amounts of proteins from protein mixtures can be analyzed.

The cross-linking reagents used in CX-MS contain reactive groups having specific reactivities for reactions with the targeted functional groups of the amino acids and possessing spacer arms of varying lengths. Supplementary Table S5 shows some typical examples of such spacer arms. If such a cross-linking reagent reacts with residues from the same protein, one can derive distance constraints between these residues. These constraints are used as input for structural modeling. If the reagent links two different proteins, one can get information about the protein interaction partners and about their quaternary structure. Since the cross-links are covalent bonds, this allows also to detect transiently and weak binding interacting proteins. If the cross-linking reaction is terminated by a quencher, then the proteins contain also cross-linker specific, so-called dead end modifications. In general one distinguishes three types of cross-links [5]: Type 0 cross-links are cross-linker related dead end modifications, also called mono-links. They can give information about the solvent accessibility of the modified residues. Type 1 cross-links are intra-linked (intra-peptide or intra-protein) cross-links (also called loop-links). From intra-protein cross-links one can derive distance constraints for the maximum distance between the two cross-linked residues inside the protein. Type 2 cross-links are inter-linked (inter-peptide or inter-protein) cross-links. Inter-protein cross-links can give information about interacting protein partners and deliver distance constraints between subunits of protein complexes.

Since the efficiency of the cross-linking reactions is low, the cross-linked peptides occur in low abundance. Therefore, an enrichment step after the enzymatic digestion, e.g. a size exclusion chromatography (SCX), is
done before the cross-linked peptides are identified [6]. Another possibility is to use trifunctional cross-linkers, which contain a handle, e.g. biotin, which can later be used to separate the cross-linked peptides by affinity chromatography [7].

The identification of the cross-linked peptides by a search engine is challenging, because the identification of the cross-links has in general quadratic computational complexity compared with the linear complexity of a database search in normal MS [8]. This is due to the fact that one has to search for all possible binary combinations of peptides. This can either be done in an exhaustive search [9] or one explicitly encodes all the cross-linked peptides in the search database [10]. Therefore cross-linkers with certain properties, which can be utilized to develop specialized and faster search strategies, were developed. Examples are CID-cleavable cross-linkers [11, 12], isotope-labelled cross-linkers [13, 14], Protein Interaction Reporters (PIR) [15], photo-cleavable cross-linkers or cross-linker with MS2 labile bonds, allowing to identify the cross-linked peptides in the following MS3 step [16]. Another category is photoactivatable cross-linkers [17-19], which are mostly not selective regarding the targeted amino acid, but allow controlling the cross-linking reaction using a UV light source. Aldehydes such as glutardialdehyde or formaldehyde are broadly specific, i.e. they react with many amino acid residues and also with DNA and can also be used for the detection of protein-DNA interactions [20].

2. Cross-linking reagents

Besides deriving distance constraints for structure determination, cross-linking is useful to study receptor-ligand interactions, for the location determination of membrane proteins and cell surface receptors, for the immobilization on solid surfaces, for the production of immunotoxins and for labelling of proteins with fluorescent dyes (Table 1). For instance, the photo-affinity labeling method gets three-dimensional information about the binding interfaces, either between two proteins or between a protein and a drug in the binding pocket of that protein receptor for a small molecule drug [21]. This method uses cross-linkers with the photo-reactive groups (azido, benzophenone and diazirines) mentioned above. These cross-linkers are able to react with C-H bonds, which is the reason for the broad non-selective reactivity of these photoactivatable cross-linkers [21, 22]. In addition, there are cross-linkers available, which can link proteins with nucleic acids, lipids or carbohydrates. For instance cross-linkers with the photo-reactive psoralen group react with pyrimidines, especially the thymines of nucleic acids [23] and diazirine reactive groups are able to react with aldehydes of sugar chains [24].

Another possibility is the site-specific incorporation of cross-linkers into a cell’s proteome by using photo-reactive amino acids analogues in order to study protein-protein interactions (PPIs) in living cells (in vivo crosslinking) [25]. One possibility is to genetically encode - by using an expanded genetic code - for the incorporation of a photo-reactive amino acid analogue like e.g. p-benzoyl-L-phenylalanine (pBpa) into the targeted protein [26-28]. For easy identification by doublets at M and M + 11, a deuterium-labelled pBpa-d11 version is used [29].
Other photo-reactive amino acids analogues that can be used are photo-L-Leu, photo-L-Ile and photo-L-Met [30] containing a photo-reactive diazirine group by which they can insert into C-H, N-H or S-H bonds, where no structural changes of the peptides due to the insertion of the diazirine ring were observed [31]. For all of them it was shown that their incorporation retains the protein structure and functions [28].

Table 1: Example use cases for cross-linking reagents usage.

| Application                                                                 | Comment                                                                                                                                                                                                 |
|----------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Deriving distance constraints for investigation of protein structure and guiding a structure-based modelling process. | Uses cross-linkers with different chemical reactivity and with different spacer lengths to derive a set of distance constraints which can restrict the search space in a subsequent structure modelling step. Often isotopically labelled heavy/light cross-linker pairs are used, since then the cross-linked peptides can be easily identified by the search software due to their characteristic doublet signals in the MS spectrum. Low concentration of protein and high cross-linker concentration favors intramolecular crosslinking. For the identification of subunits, cleavable cross-linkers are useful, so that by 2D-gel electrophoresis the subunits of the complex can be detected. |
| Study of protein-protein or protein-DNA and receptor-ligand interactions for the investigation of signal transduction. | Uses protein cross-linkers to establish an association between two interacting proteins resp. a receptor protein and its ligand for the derivation of the quaternary structure of protein complexes [32].                                                                                     |
| Determination of the location of membrane proteins and identification of cell surface receptors. | Integral membrane proteins react with water-insoluble cross-linkers, but not with hydrophilic cross-linkers. Membrane impermeable cross-linkers like Sulfo-NHS esters and pegylated cross-linkers, due to their hydrophilic nature, react only with proteins located on the cell surface. |
| Immobilization of proteins on solid supports. | Typical supports are e.g. nitrocellulose membranes, glass, agarose, polystyrene plates or beads. Allows e.g. to produce protein nanosurfaces for the production of biosensors or analytical devices like e.g. protein chips [33]. |
| Use as immunological tools, e.g. for the production of immunotoxins or the coupling of haptens to carrier proteins for getting antigens for antibody production. | Linking toxins to antibodies for getting molecules that can be used for targeted killing of cancer cells. Often cross-linkers with a cleavable disulfide S-S bond are used, because most cells are able to cleave these and thereby the bound toxin is released. |
| Transfer of labels or reactive groups | Cross-linkers can also be used to modify target groups on a |
protein or to add spacers for later coupling reactions.

| Labeling of proteins. | Enable identification, enrichment, purification or detection of proteins by transferring labels like e.g. biotin or fluorescent dyes from a known bait protein to an unknown interacting prey protein. |
|-----------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| In vivo crosslinking [34]. | Stabilizing transient endogenous protein-protein or receptor-ligand interactions. Photo-reactive amino acids analogues are used to incorporate the cross-linkers into the cell’s proteome. |

3. Properties and selection criteria for choosing cross-linking reagents

Criteria for the selection of a cross-linker were defined by Jin Lee [35] and Back et.al. [36]. The most important criteria are the specificity, which is given by the reactive groups, and the spacer lengths, especially with respect to getting distance constraints for later protein structure modeling.

Further possible criteria, e.g. the ability to permeate membranes, determine which kinds of proteins can be cross-linked. For an easier detectability of the cross-linked peptides by the used search machines, the availability of isotopically coded versions of the cross-linker or the cleavability of the cross-linker can be important, since then such cross-linked peptides are more easily detectable by the used search engines. When such cross-linkers are not available, one can make use of trifunctional cross-linkers with an additional handle, which allow for an additional enrichment step before the mass spectrometry analysis. Another possibility is the enrichment by using size exclusion chromatography [37]. Table 2 gives an overview about cross-linker properties, documented in the xmod ontology and useful for cross-linker selection. In order to expand the applicability of the xmod ontology, we included also some properties, which are not specific for XL-MS. Some vendors also offer cross-linker selection guides and/or interactive selection tools for choosing proper cross-linkers (see Supplementary Table S3).

Table 2: Cross-linker properties

| Property               | Comment / Reason / Example                                                                                                                                                                                                 |
|------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Chemical specificity   | The specificity is determined by the reactive groups, targeting specific groups of the different amino acids.                                                                                                           |
| Reaction conditions    | e.g. buffer system, pH, concentrations, …                                                                                                                                                                                    |
| Spacer length          | Distance between the cross-linked proteins acting as molecular rulers for deriving spatial constraints for protein structure modelling [38].                                                                              |
| Spacer composition     | Determines hydrophobicity. For instance, a spacer containing PEG (polyethyleneglycol) makes the cross-linker hydrophobic and therefore prevents hydrophobic aggregation.                                                     |
If the cross-linker can be cleaved, either chemically, photo-activated or by fragmentation, e.g. CID (Collision Induced Dissociation). This cleavability e.g. allows the release of labeled cross-linked peptides from the supports used in an affinity based purification step [39].

Determines if the cross-linker can detect cell surface proteins or can cross-link integral membrane proteins (which are hydrophobic) resp. if a cross-linker can permeate into cells (e.g. for in vivo cross-linking).

Makes the cross-link identification process by the search program easier, e.g. by searching for the characteristic doublets [29] of cross-links containing the heavy resp. light version of the cross-linker.

If spontaneously reactive with the targeted functional groups of residues or photo-reactive (non-selective), e.g. for two-step cross-linking reactions.

Using cross-linkers with a handle (e.g. biotin or a fluorophore) for easy enrichment of the cross-linked peptides.

4. Derivatization reagents
Because derivatization reagents like cross-linker have functional groups specifically reacting with chemical target groups, their description is very similar to that of the cross-linker reagents. Therefore, we decided to include them into the XLMOD ontology, so that beside the cross-linker reagents used in proteomics we also included derivatization reagents, useful for annotation in metabolomics and lipidomics. These derivatization reagents are either used in GC/MS for enhancing the volatility of polar metabolites, allowing them to pass easily into the gas phase [40, 41]. Similarly in LC-MS approaches the derivatization is done in order to enhance the LC signal intensities, the ionization in the ESI, (ElectroSpray Ionization) and/or the MS-detectability of the metabolite by a more efficient fragmentation [42]. The used derivatization reagents should specifically react with the targeted metabolites producing a high yield of the derivative [42]. The main groups of derivatization reagents for GC-MS are acylation, alkylation and silylation reagents, which replace active hydrogens (e.g. from -SH, -OH, -NH, -COOH groups) by an acyl, alkyl resp. silyl group, which confers the derivatized target molecule a more hydrophobic character, which again enhances their volatility. Derivatization reagents for LC-MS at one hand have a functional group for the formation of the derivative, and on the other hand they introduce a group, which allows an easier detection of the derivative. This improved detectability is accomplished either in the LC part by introducing a chromophore, a fluorophore or an electrophore group for detection by UV/VIS light, by fluorescence or by electrochemical detectors [43]. Another possibility is to enhance the detectability in the MS part by improved ion formation in the ESI or by making the fragmentation more efficient. In mzTab-M [44] the used derivatization reagents should be
reported in the metadata section under derivatization_agent[1-n] as specified by the mzTab-M specification
document available at https://github.com/HUPO-PSI/mzTab.

5. xlmod ontology for cross-linking and derivatization reagents

Although there are several articles listing cross-linkers for use in structural proteomics [45-47], there was still
an urgent need for a structured ontology, which can serve both as a database for choosing proper
cross-linkers, as well as source for uniquely describing identified cross-linked peptides and peptides with
post-translational modifications derived from cross-linkers.
For building up the xlmod ontology we scanned vendor websites, publications and catalogues (see
Supplementary Tables S1 - S3), relevant papers from the PubMed literature database [48] and some
relevant books [49, 50].

The most important feature for the selection of a cross-linker reagent is its specificity for amino acids [51]
and then it’s spacer length. The amino acid residues, which are currently targetable by cross-linkers, are
summarized in Supplementary Table S4.
It was shown, that the availability of cross-linking reagents targeting the acidic carboxyl groups of Asp and
Glu allows deriving distance constraints that orthogonally complement constraints obtained from Lys
cross-linking [51, 52]. It can be expected that cross-linkers specifically targeting other amino acids, e.g. the
indole amine of tryptophan, can further enhance the power of the protein cross-linking method as a 3D
structure determination tool.

The reactive groups occurring in cross-linkers and their amino acid specificities are reviewed in [25, 46, 47,
53], and summarized in Table 3.

| Reactive group                  | Target functional group | Specificity for amino acids                     | Photo-reactive | Reference (PMID) |
|--------------------------------|-------------------------|------------------------------------------------|----------------|-----------------|
| formaldehyde                   | primarily amines,       | K,N,R,Q,Y, N-term secondary: C,F,H,S,T,W       |                | 25979347        |
|                               | but broadly specific    |                                                |                |                 |
| 1-hydroxy-7-azabenzotriazole   | amines                  | K,N,Q,R,Protein N-term                         |                | 19994840        |
| ole                            |                         |                                                |                |                 |
| amine                          | carboxyl groups         | D,E                                            |                |                 |
| aryl halide                    | primary amines          | K,N,Q,R,Protein N-term                         |                |                 |
| aryl azides (e.g. phenyl-,     | insertion into C-H      | nonselective                                   | yes            | 22641729        |
| nitrophenyl-)                  | and N-H bonds;          |                                                |                |                 |
| Compound                                | Reactivity | Selectivity | Response |
|-----------------------------------------|------------|-------------|----------|
| hydroxyphenyl-, tetrafluorophenyl-, perfluoroaryl azide | primary amines | --- | nonselective | yes |
| azido-methylcoumarin                  | ---        | nonselective | yes |
| benzopheneone                          | ---        | nonselective | yes |
| carbodiimide                           | carboxyl-to-amine reactive | (K,N,Q,R,Protein N-term)&(E,D,Protein in C-term) | --- |
| carbonyl                               | hydrazines | ---         | --- |
| carboxyl                               | amines     | K,N,Q,R,Protein N-term | --- |
| diazirine                              | ---        | nonselective | yes | 21778064 |
| diazogroup                             | ---        | nonselective | yes |
| dimethylether                          | carboxyls  | D,E         | --- |
| glyxal                                 | arginine-specific | R | --- |
| haloaceyl (chloro-, bromo-, iodoacetyl) | sulphydryls | C | --- |
| hydrazide                              | carboxyls  | D,E,Protein C-term | --- |
| hydrazone                              | carboxyls  | D,E,Protein C-term | --- |
| hydroxybenzotriazole                   | amines     | K,N,Q,R,Protein N-term | 19994840 |
| hydroxymethyl phosphine                | amines     | K,N,Q,R,Protein N-term | --- |
| imidoester                             | amines     | K,N,Q,R,Protein N-term | --- |
| isocyanate                             | hydroxyl   | S,T,Y        | --- |
| maleimide                              | sulphydryl | ---         | --- |
| methanethiosulfonate                   | sulphydryl | ---         | --- |
| N-hydroxyphthalimide                   | amines     | K,N,Q,R,Protein N-term | 19994840 |
| NHS (N-hydroxysuccinimide) ester       | lysines and N-termini, but also hydroxyls | K,Protein N-term secondary: S,T,Y | --- |
| pentafluorophenyl                      | amines     | K,N,Q,R,Protein N-term | 22067100 |
| platinum (II)                          | thioether, sulphydryl, and | C,H,M | 21591778 |
| Term                     | Def                      |
|-------------------------|--------------------------|
| imidazole               |                         |
| psoralen                | pyrimidines, esp.        |
|                         | thymins                 |
| pyridinyl disulfide     | sulfhydryls             |
| (pyridyldithiol)        |                         |
| S-acetyl                | incorporates            |
|                         | sulfhydryls             |
| Sulfo NHS               | lysines and             |
| (N-hydroxysulfosuccinimide) ester | N-termini, but also    |
|                         | hydroxyls               |
| thioimidate             | amines                  |
| vinyl sulfone           | amines, sulfhydryls,   |
|                         | hydroxyls               |

It must be mentioned here that the "specificities" for amino acids are not absolute, but rather describe a tendency for preferred reaction targets, because the specificities can vary with the reaction conditions [54]. For instance NHS esters and Sulfo NHS esters react primarily with the amine group of lysines and with protein N-terminals, but to a lower degree they can also react with the hydroxyl groups of serine, threonine and tyrosine [25, 55], as specified by the property "secondarySpecificities".

In general, each cross-linking reagent term has a unique and fixed id, a short name, a definition, an optional list of synonyms and a list of properties with their values, the specification of its parent terms and then a list of relationships, describing e.g. the chemical reactivity.

As a concrete example, the entry for the reactive group "NHS ester" is defined as follows:

```
[Term]
  id: xmod:00101
  name: NHS ester
  def: "A reactive group (N-hydroxysuccinimide) that reacts with lysines and N-terminals but also serines, threonines and tyrosines." [PSI:XL]
  synonym: "N-hydroxysuccinimide ester" EXACT []
  property_value: specificities: "(K,Protein N-term)" xsd:string
  property_value: secondarySpecificities: "(S,T,Y)" xsd:string
  is_a: xmod:00100 ! reactive group
  relationship: has_property xmod:00013 ! membrane permeable
  relationship: is_reactive_with xmod:00028 ! primary amine reactive
```
6. General structure of the xlmod ontology

The xlmod ontology consists of four main branches of cross-linker or derivatization reagent terms and defines several relations, which can be used to describe such terms in general (see Figure 1). The three main branches are:

- cross-linker related chemical modification
- cross-linking entity
- derivatization entity
- label transfer reagent
The cross-linker related chemical modification branch contains partially reacted products of cross-linkers, e.g. modifications resulting from dead-ends (mono-links), which are no real cross-links, since they do not link two chains of proteins. In principle one can quench the cross-linking reaction with any amine, but the mostly ammonium salts or Tris are used. Therefore mainly amidated or hydrolyzed cross-linkers, but also thioesterized or alkenized forms are listed in xlmod. The amidated modifications are subdivided into ammonium amidated and Tris amidated forms, since these differ slightly in the resulting modifications and their masses, because ammonium salts convert to \(-\text{C(=O)}-\text{NH}_2\), whereas the use of Tris(hydroxymethyl)-aminomethane as a quenching substance leads to the formation of dead ends of the form \(-\text{C(=O)}-\text{NHC(CH}_2\text{OH)}_3\). At the moment we have no instances of Tris amidated forms, but we distinguished these cases to allow them to be added later. The label transfer reagents are either trifunctional cross-linkers containing a handle, e.g. biotin, for enrichment or for transferring a label to a prey protein.

The main branches under cross-linking entity are:

- amino acid and their derivatives (containing photo-reactive amino acid analogs)
- the cross-linkers in the narrower sense (resp. the cross-linked products derived from them)
- cross-linking attributes for a more detailed characterization of the cross-linkers, e.g. their cleavability, and cross-linker derived chemical modifications
- handles used for detection, affinity enrichment or purification of cross-linked peptides (see Supplementary Table S6)
- reactive group attributes for a more detailed description of reactive groups

Important cross-linking attributes are the reactive groups, which have specific chemical reactivity towards functional target groups, and which are listed under the reactive group attributes.

The branch derivatization entity contains sub-branches for

- chromophores (for UV/VIS detection)
- fluorophores (for fluorescence detection)
- electrophores (for electrochemical detection)
- GCMS derivatization reagents
- LCMS derivatization reagents.

A typical example for a cross-linker term is:

```plaintext
[Term]
  id: xlmod:02001
  name: DSS
  def: "Disuccinimidyl suberate." [CAS:68528-80-3, PubChem_Compound:100658, PMID:16944939, ChemSpiderID:90944, ChemicalBookNo:CB1219157]
  synonym: "Bis(succinimidyl) suberate" EXACT []
  synonym: "Suberic acid bis(N-hydroxysuccinimide ester)" EXACT []
  synonym: "Disuccinimidyl octanedioate" EXACT []
```
The cross-linker entries possess a unique ID and (short) name. Then follows a definition line containing the long name as well as a list of database cross references, which can be of one of the types listed in Table 4. The PMID cross references and web links can occur several times for a given definition, whereas the other possible cross references must be unique.

Table 4: Used database cross references (dbxref’s) in the xlmod.obo ontology

| DB cross reference | DB name                          | Link                                                                 |
|-------------------|----------------------------------|---------------------------------------------------------------------|
| Beilstein         | Beilstein Registry Number        | https://www.elsevier.com/solutions/reaxys                          |
| CAS               | Chemical Abstract Service        | https://www.cas.org/content/chemical-substances                     |
| CBNumber          | Chemical Book                    | http://www.chemicalbook.com/Search_EN.aspx                         |
| ChemSpiderID      | ChemSpider                       | http://www.chemspider.com                                         |
| MDL               | former Molecular Design Ltd., now BIOVIA | http://accelrys.com                                                              |
| PMID              | Pubmed ID                        | http://www.ncbi.nlm.nih.gov/pubmed/advanced                        |
| PubChem_Compound  | PubChem                          | https://www.ncbi.nlm.nih.gov/pccompound/advanced                    |
| http://...        | Web link                         | specific web link, where the cross-linker is defined                  |
| DOI:...           | Web link                         | web link, e.g. for publications not indexed in PubMed              |

The chemical substance registry references can be used, if one wants to get more detailed information about a cross-linker, e.g. the IUPAC name, a full list of synonyms, the SMILES string, InChI key, 3D formula, and so on. From the Chemical Book repository, one can get also MDL .mol files containing the exact coordinates of the 3D structure. In principle one can then calculate the spacer arm lengths from the coordinates in these mol files [56], if one is acquainted with the reaction mechanism, so that one exactly knows which atoms belong to the spacer chain of the resulting cross-linked peptides. Details of that format are described at http://accelrys.com/products/collaborative-science/biovia-draw/ctfile-no-fee.html (accessed 08/2018) and http://media.accelrys.com/downloads/ctfile-formats/ctfile-formats.zip. Other possibilities to get the spacer arm length are to calculate them from force filed simulations [38] or to calculate
them from 3D coordinates. If the coordinates are unknown, one can calculate them from the SMILES strings, which one can get from PubChem, by using cheminformatics programs like e.g. the 3D structure generator CORINA, available from http://www2.chemie.uni-erlangen.de/services/telespec/corina/ (accessed 08/2018) [57], which converts the SMILES string into 3D coordinates. Another possibility is to measure the spacer arm lengths with chemical drawing programs like e.g. BIOVIA Draw (http://accelrys.com/resource-center/downloads/freeware/, accessed 08/2018). A tool that can be used to calculate the m/z ratio of cross-linked peptides is the computer program GPMAW [58] (http://www.gpmaw.com/html/cross-linking.html and http://www.gpmaw.com/html/new_cross-linker.html, accessed 08/2018).

Further available software tools for cross-linking analysis and identification are already summarized by Mayne and Patterton [59], Sinz [60], Leitner et.al. [16], Tran et.al. [32], and in the list of cross-linking software tools at https://omictools.com/cross-linked-peptide-identification-category (accessed 08/2018).

In the xlmod ontology we only listed synonyms, which are commonly used by the cross-linking community. The bridge formula (sum formula of the spacer arm) is given in Hill notation [61]. In addition, the monoisotopic mass calculated from the bridge formula, is specified. Then the number of reaction sides is given and the cross-linker is either classified as homofunctional (i.e. the cross-linker has only one sort of reactive group) or heterofunctional (i.e. containing more than one different reactive groups). In the DSS example above the number of reaction sites is 2, so that it’s a homobifunctional cross-linker. At the end the type of reactive groups are listed.

In order to avoid redundancy, the specificities for the amino acids are not repeated in the term for the cross-linker, so that the specificities must be looked up at the terms for the listed reactive groups of that cross-linker. This means that the properties from reactive groups are inherited and apply to the cross-linking reagents containing the respective reactive group. For example, it follows that a cross-linker with a NHS ester reactive tends to be more membrane permeable, whereas one with a Sulfo-NHS ester reactive group is more hydrophilic. The overall hydrophobicity is also influenced by the hydrophilicity of the spacers and affinity arms, e.g. PEG spacers make a reagent the more hydrophilic the longer the PEG spacer group is. Likewise, it is clear, that a cross-linker with a pyridinyldisulfide reactive group possesses a cleavable S-S bond or that a cross-linker with a biotin handle is enrichable.

An example for a cross-linker related post-translational modification is the term for “hydrolyzed DSS”, where the relationship “is_side-product_of” specifies from what cross-linker reagent the chemical modification is derived from:

[Term]

id: xlmol:01002
name: hydrolyzed DSS
def: "Hydrolyzed disuccinimidyl suberate." [PSI:XL]

property_value: deadEndFormula: "C8 H12 O3" xsd:string
property_value: monoisotopicMass: "156.07864431" xsd:double
is_a: xmod:00096 ! hydrolization cross-linker related chemical modification
relationship: is_partially_reacted xmod:00011 ! hydrolyzed
relationship: is_side_product_of xmod:02001 ! DSS

The following is an example for a labelled cross-linker with the relationship "is_labelled":

[Term]
id: xmod:02002
name: DSS-d4
def: "Deuterium labelled disuccinimidyl 2,2,7,7-suberate." [PubChem_Compound:91757798, PMID:11354472]
property_value: bridgeFormula: "C8 D4 H6 O2" xsd:string
property_value: doubletDeltaMass: "4.02508" xsd:double
property_value: monoisotopicMass: "142.093186586" xsd:double
property_value: reactionSites: "2" xsd:nonNegativeInteger
property_value: spacerLength: "11.4" xsd:float
is_a: xmod:00005 ! homofunctional cross-linker
relationship: has_reactive_group xmod:00101 ! NHS ester
relationship: is_labelled xmod:00010 ! deuterium labelled

The property "doubletDeltaMass" gives the m/z difference between the labelled and the unlabelled versions of the cross-linker and can be used by cross-linking search-engines for easy detection of the corresponding cross-linked peptides.

As an example for an entry of a photo-reactive amino acid analogue the entry for pBpa-d11 is specified as:

[Term]
id: xmod:01904
name: pBpa-d11
def: "Deuterium labelled p-benzoyl-L-phenylalanine." [PMID:18704231]
comment: A photoreactive deuterium-labelled amino acid analog of L-Phenylalanine for incorporation during protein synthesis that can be used for in vivo labelling, cross-linking and protein-protein interaction studies in live cells.
property_value: reactionSites: "1" xsd:nonNegativeInteger
is_a: xmod:00049 ! amino acid derivative
Since the PSI-MOD ontology is not maintained anymore, we also added a special branch to the xlm0d ontology that contains label transfer reagents, which are primarily used for transferring labels and not for cross-linking. An example would be

[Term]
id: xlm0d:01800
name: Psoralen-PEG3-Biotin
def: "Psoralen-PEG3-Biotin biotinylation reagent." [PSI:XL]
property_value: hydrophilicPEGchain: "3" xsd:nonNegativeInteger
property_value: reactionSites: "1" xsd:nonNegativeInteger
property_value: spacerLength: "36.86" xsd:float
is_a: xlm0d:00003 ! label transfer reagent
relationship: has_handle xlm0d:00051 ! biotin
relationship: has_property xlm0d:00014 ! hydrophilic
relationship: has_reactive_group xlm0d:00134 ! psoralen

At the moment the xlm0d ontology contains 16 type definitions and 1083 terms, of which 248 are bifunctional cross-linker terms, 103 terms belong to the class of cross-linker related chemical modifications and 84 terms describe chemical reactivities, reactive groups and reactive group attributes. In addition there are 174 terms for GC/MS derivatization reagents and 324 terms for LC/MS derivatization reagents. The other terms describe zero-length and trifunctional cross-linkers, label transfer reagents, amino acids and their derivatives, handles, chemical reactivities and other attributes and properties of the cross-linkers. For details see the Supplementary Table S7 and S8.

7. Reporting of cross-linking results
Controlled vocabulary (CV) terms from the xlm0d ontology and from the psi-ms ontology [62] are used together for documenting cross-linking searches in version 1.2 of the XML-based mzIdentML standard format [63], see Figure 2. For that one has to specify the CV term “crosslinking search” (MS:1002494) in the <SpectrumIdentificationProtocol> element of the mzIdentML file [63]. The two cross-linked peptides must have <Modification> elements, which are marked as cross-linked peptides by using the CV terms “cross-link donor” and “cross-link acceptor”, which are defined in the psi-ms ontology [62]. These two corresponding peptides are matched together by having the same value. The mzIdentML validator [64] checks that these
two CV terms, grouped together by the same value, are located under different <Modification> elements. By convention, the donor peptide is the one with the longer peptide sequence. In case that both peptides have the same length, the donor is defined as the one, which comes first in the alphabetical order of the peptide sequence.

By convention, the cross-link acceptors are reported by a <Modification> element with a monoisotopic mass delta of zero, and the cross-link donor with the actual mass of the cross-link. In addition, the peptide, which is the "cross-link donor" must have another CV term annotating its <Modification> element [63]. This term must be a term located under the branch "cross-linker" (xlmod:00004) from the xlmod ontology, and describes the cross-linking reagent. In contrast, a dead-end modification is annotated by only one CV term, referencing a child term of "cross-linker related chemical modification" (xlmod:00002) from the xlmod ontology.

Since there are always pairs of donor and receiver peptides, there must be two <PeptideEvidenceRef> elements (referencing the cross-linked peptides), that occur paired with the same charge state in two <SpectrumIdentificationItem> elements under the same <SpectrumIdentificationResult> element with the same value for the CV term "cross-link spectrum identification item" (MS:1002511) annotating these <SpectrumIdentificationItem> elements. If one uses isotope-labelled cross-linker reagents, there are even four <SpectrumIdentificationItem> elements for the same cross-link and charge state, since donor and acceptor can occur now cross-linked with the heavy and light version of the used cross-linker reagent. This rule is also checked by the mzIdentML validator [64].
We want to mention here, that beside mzIdentML 1.2 there is also an alternative open data format for reporting cross-linking results, developed by Hoopmann et.al. [65] and is an extension of the pepXML format [66].

8. Outlook to the further development of xlmod

Then HUPO-PSI intends to steadily further develop the xlmod ontology. Currently, on the one hand there are still some cross-linkers missing and on the other hand for some cross-linkers there are still some data missing. Our hope is that the cross-linking community will contribute such missing data and request terms for still missing respective newly developed cross-linkers via the ‘psidev-ms-vocab’ mailing list at https://lists.sourceforge.net/lists/listinfo/psidev-ms-vocab and thereby contribute to make this ontology a standard source for cross-linkers, their related chemical modifications, and derivatization reagents. Then this ontology can be utilized by specialized cross-linking identification and/or analysis software like e.g. xiFDR [67], SIM-XL [68] or OpenMS [69], which allow to report their results in the mzIdentML 1.2 [63] standard format. The xlmod ontology is also open for bio-conjugative reagents used in cross-linking of either RNA-RNA [70], or proteins to other biomolecules like e.g. DNA / RNA [23, 71-73], lipids [74, 75] or sugars [76-78]. For instance there are lipid analogues defined, which are useful for describing the interaction of lipids with membrane proteins [75], the terms for nucleobases and their derivatives can be used for the description of protein-RNA interactions [72] and the sialic acid analogue terms are intended for the description of glycan-protein crosslinking [78] experiments.

Conceivable is the use of xlmod as a database for the development of an open-source tool for the planning of cross-linking experiments, allowing the selection of the most promising cross-linking reagents dependent of the detailed aims of the planned experiment. For the easy access to xlmod the development of a general Java access library, resolving all the inheritance dependencies between the different terms, e.g. between the reactive groups and the cross-linking reagents, would also be desirable.

Desirable for the reproducible application of CX-MS in integrated structure determination and prediction approaches are also standardized experimental protocols, since the specificities of the cross-linking reagents depends on the reaction conditions [79], as well as robust, standardized and automatable data analysis workflows [80, 81] for such integrated structure determination and modeling methods. The xlmod ontology is one building block for such a future highly automated high-throughput application of CX-MS in structural biology.

Acknowledgements:
Gerhard Mayer was funded by the BMBF grant de.NBI - German Network for Bioinformatics Infrastructure (FKZ 031 A 534A).
References:

1. Benesch JL, Ruotolo BT. Mass spectrometry: come of age for structural and dynamical biology, Curr Opin Struct Biol 2011;21:641-649.
2. Fabris D, Yu ET. Elucidating the higher-order structure of biopolymers by structural probing and mass spectrometry: MS3D, Journal of Mass Spectrometry 2010;45:841-860.
3. Holding AN. XL-MS: Protein cross-linking coupled with mass spectrometry, Methods 2015;89:54-63.
4. Leitner A, Faini M, Stengel F et al. Crosslinking and Mass Spectrometry: An Integrated Technology to Understand the Structure and Function of Molecular Machines, Trends in Biochemical Sciences 2016;41:20-32.
5. Schilling B, Row RH, Gibson BW et al. MS2Assign, automated assignment and nomenclature of tandem mass spectra of chemically crosslinked peptides, J Am Soc Mass Spectrom 2003;14:834-850.
6. Fritzsche R, Ihling CH, Gotze M et al. Optimizing the enrichment of cross-linked products for mass spectrometric protein analysis, Rapid Commun Mass Spectrom 2012;26:653-658.
7. Sinz A, Kalkhof S, Ihling C. Mapping protein interfaces by a trifunctional cross-linker combined with MALDI-TOF and ESI-FTICR mass spectrometry, J Am Soc Mass Spectrom 2005;16:1921-1931.
8. Yang L, Tang X, Weisbrod CR et al. A photocleavable and mass spectrometry identifiable cross-linker for protein interaction studies, Analytical Chemistry 2010;82:3556-3566.
9. Yu FC, Li N, Yu WC. ECL: an exhaustive search tool for the identification of cross-linked peptides using whole database, BMC Bioinformatics 2016;17.
10. Yilmaz S, Drepper F, Hulstaert N et al. Xilmass: A New Approach toward the Identification of Cross-Linked Peptides, Analytical Chemistry 2016;88:9949-9957.
11. Gotze M, Pettelkau J, Fritzsche R et al. Automated assignment of MS/MS cleavable cross-links in protein 3D-structure analysis, J Am Soc Mass Spectrom 2015;26:83-97.
12. Sinz A. Divide and conquer: cleavable cross-linkers to study protein conformation and protein-protein interactions, Anal Bioanal Chem 2016.
13. Petrotchenko EV, Borchers CH. ICC-CLASS: isotopically-coded cleavable crosslinking analysis software suite, BMC Bioinformatics 2010;11:64.
14. Rinner O, Seebacher J, Walzthoeni T et al. Identification of cross-linked peptides from large sequence databases, Nat Methods 2008;5:315-318.
15. Tang X, Bruce JE. A new cross-linking strategy: protein interaction reporter (PIR) technology for protein-protein interaction studies, Mol Biosyst 2010;6:939-947.
16. Leitner A, Walzthoeni T, Kahraman A et al. Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics, Mol Cell Proteomics 2010;9:1634-1649.
17. Preston GW, Wilson AJ. Photo-induced covalent cross-linking for the analysis of biomolecular interactions, Chem Soc Rev 2013;42:3289-3301.
18. Dubinsky L, Krom BP, Meijler MM. Diazirine based photoaffinity labeling, Bioorg Med Chem 2012;20:554-570.
19. Das J. Aliphatic Diazirines as Photoaffinity Probes for Proteins: Recent Developments, Chem Rev 2011;111:4405-4417.
20. Lu K, Ye WJ, Zhou L et al. Structural Characterization of Formaldehyde-Induced Cross-Links Between Amino Acids and Deoxynucleosides and Their Oligomers, J Am Chem Soc 2010;132:3388-3399.
21. Robinette D, Neamati N, Tomer KB et al. Photoaffinity labeling combined with mass spectrometric approaches as a tool for structural proteomics, Expert Rev Proteomics 2006;3:399-408.
22. Hatanaka Y. Development and leading-edge application of innovative photoaffinity labeling, Chem Pharm Bull (Tokyo) 2015;63:1-12.
23. Sastry SS, Ross BM, P’Arraga A. Cross-linking of DNA-binding proteins to DNA with psoralen and psoralen furan-side monoadducts. Comparison of action spectra with DNA-DNA cross-linking, J Biol Chem 1997;272:3715-3723.
24. Tomohiro T, Hashimoto M, Hatanaka Y. Cross-linking chemistry and biology: development of multifunctional photoaffinity probes, Chem Rec 2005;5:385-395.
25. Sinz A. Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions, Mass Spectrom Rev 2006;25:663-682.
26. Hino N, Okazaki Y, Kobayashi T et al. Protein photo-cross-linking in mammalian cells by site-specific incorporation of a photoreactive amino acid, Nat Methods 2005;2:201-206.
27. Farrell IS, Toroney R, Hazen JL et al. Photo-cross-linking interacting proteins with a genetically encoded benzophenone, Nat Methods 2005;2:377-384.
28. Lee HS, Dimla RD, Schultz PG. Protein-DNA photo-crosslinking with a genetically encoded benzophenone-containing amino acid, Bioorganic & Medicinal Chemistry Letters 2009;19:5222-5224.
29. Wilkins BJ, Daggett KA, Cropp TA. Peptide mass fingerprinting using isotopically encoded photo-crosslinking amino acids, Molecular Biosystems 2008;4:934-936.
30. Lossl P, Sinz A. Combining Amine-Reactive Cross-Linkers and Photo-Reactive Amino Acids for 3D-Structure Analysis of Proteins and Protein Complexes, Methods Mol Biol 2016;1394:109-127.
31. Kolbel K, Ihling CH, Sinz A. Analysis of peptide secondary structures by photoactivatable amino acid analogues, Angew Chem Int Ed Engl 2012;51:12602-12605.
32. Tran BQ, Goodlett DR, Goo YA. Advances in protein complex analysis by chemical cross-linking coupled with mass spectrometry (CXMS) and bioinformatics, Biochim Biophys Acta 2016;1864:123-129.
33. Bhakta SA, Evans E, Benavidez TE et al. Protein adsorption onto nanomaterials for the development of biosensors and analytical devices: a review, Anal Chim Acta 2015;872:7-25.
34. Kaake RM, Wang X, Burke A et al. A new in vivo cross-linking mass spectrometry platform to define protein-protein interactions in living cells, Mol Cell Proteomics 2014;13:3533-3543.
35. Jin Lee Y. Mass spectrometric analysis of cross-linking sites for the structure of proteins and protein complexes, Mol Biosyst 2008;4:816-823.
36. Back JW, de Jong L, Muijsers AO et al. Chemical cross-linking and mass spectrometry for protein structural modeling, J Mol Biol 2003;331:303-313.
37. Leitner A, Reischl R, Walzthoeni T et al. Expanding the Chemical Cross-Linking Toolbox by the Use of Multiple Proteases and Enrichment by Size Exclusion Chromatography, Molecular & Cellular Proteomics 2012;11.
38. Green NS, Reisler E, Houk KN. Quantitative evaluation of the lengths of homobifunctional protein cross-linking reagents used as molecular rulers, Protein Science 2001;10:1293-1304.
39. Rudolf GC, Heydenreuter W, Sieber SA. Chemical proteomics: ligation and cleavage of protein modifications, Curr Opin Chem Biol 2013;17:110-117.
40. Danielson ND, Gallagher PA, Bao J. Chemical Reagents and Derivatization Procedures in Drug Analysis. Chichester: John Wiley & Sons Ltd, 2000.
41. Orata F. Derivatization Reactions and Reagents for Gas Chromatography Analysis. IntechOpen, 2011.
42. Danielson ND, Gallagher PA, Bao J. Chemical Reagents and Derivatization Procedures in Drug Analysis. Chichester: John Wiley & Sons Ltd, 2000.
43. Poole CF. Derivatization in Liquid Chromatography. Elsevier, 2017.
44. Hoffmann N, Rein J, Sachsenberg T et al. mZTab-M: A Data Standard for Sharing Quantitative Results in Mass Spectrometry Metabolomics, Anal Chem 2019;91:3302-3310.
45. Mattson G, Conklin E, Desai S et al. A practical approach to crosslinking, Mol Biol Rep 1993;17:167-183.
46. Serpa JJ, Parker CE, Petrotchenko EV et al. Mass spectrometry-based structural proteomics, Eur J Mass Spectrom (Chichester, Eng) 2012;18:251-267.
47. Paramelle D, Miralles G, Subra G et al. Chemical cross-linkers for protein structure studies by mass spectrometry, Proteomics 2013;13:438-456.
48. Lu Z. PubMed and beyond: a survey of web tools for searching biomedical literature, Database (Oxford) 2011;2011:baq036.
49. Hermanson GT. Bioconjugate Techniques. Oxford: Elsevier, 2013.
50. Wong SS, Jameson, D.M. Chemistry of Protein and Nucleic Acid Cross-Linking and Conjugation. CRC Press Inc, 2011.
51. Hofmann T, Fischer AW, Meiler J et al. Protein structure prediction guided by crosslinking restraints - A systematic evaluation of the impact of the crosslinking spacer length, Methods 2015;89:79-90.
52. Leitner A, Joachimiak LA, Unverdorben P et al. Chemical cross-linking/mass spectrometry targeting acidic residues in proteins and protein complexes, Proc Natl Acad Sci U S A 2014;111:9455-9460.
53. Zybailov BL, Glazko GV, Jaiswal M et al. Large Scale Chemical Cross-linking Mass Spectrometry Perspectives, J Proteomics Bioinform 2013;6:001.
54. Kalkhof S, Sinz A. Chances and pitfalls of chemical cross-linking with amine-reactive N-hydroxysuccinimide esters, Anal Bioanal Chem 2008;392:305-312.
55. Madler S, Bich C, Touboul D et al. Chemical cross-linking with NHS esters: a systematic study on amino acid reactivities, J Mass Spectrom 2009;44:694-706.
56. Dalby A, Nourse JG, Hounshell WD et al. Description of Several Chemical-Structure File Formats Used by Computer-Programs Developed at Molecular Design Limited, Journal of Chemical Information and Computer Sciences 1992;32:244-255.
57. Sadowski J, Gasteiger J, Klebe G. Comparison of Automatic Three-Dimensional Model Builders Using 639 X-ray Structures, Journal of Chemical Information and Modeling 1994;34:1000-1008.
58. Peri S, Steen H, Pandey A. GPMAW--a software tool for analyzing proteins and peptides, Trends in Biochemical Sciences 2001;26:687-689.
59. Mayne SL, Patterton HG. Bioinformatics tools for the structural elucidation of multi-subunit protein complexes by mass spectrometric analysis of protein-protein cross-links, Brief Bioinform 2011;12:660-671.
60. Sinz A, Arlt C, Chorev D et al. Chemical cross-linking and native mass spectrometry: A fruitful combination for structural biology, Protein Sci 2015;24:1193-1209.
61. Hill EA. ON A SYSTEM OF INDEXING CHEMICAL LITERATURE; ADOPTED BY THE CLASSIFICATION DIVISION OF THE U. S. PATENT OFFICE. J. Am. Chem. Soc., 1900, 478–494.
62. Mayer G, Montecchi-Palazzi L, Ovelleiro D et al. The HUPO proteomics standards initiative- mass spectrometry controlled vocabulary, Database (Oxford) 2013;2013:bat009.
63. Vizcaíno JA, Mayer G, Perkins S et al. The mzIdentiML Data Standard Version 1.2, Supporting Advances in Proteome Informatics, Mol Cell Proteomics 2017;16:1275-1285.
64. Ghali F, Krishna R, Lukasse P et al. Tools (Viewer, Library and Validator) that facilitate use of the peptide and protein identification standard format, termed mzIdentML, Mol Cell Proteomics 2013;12:3026-3035.
65. Hoopmann MR, Mendoza L, Deutsch EW et al. An Open Data Format for Visualization and Analysis of Cross-Linked Mass Spectrometry Results, J Am Soc Mass Spectrom 2016.
66. Keller A, Eng J, Zhang N et al. A uniform proteomics MS/MS analysis platform utilizing open XML file formats, Mol Syst Biol 2005;1:2005 0017.
67. Fischer L, Rappsilber J. Quirks of Error Estimation in Cross-Linking/Mass Spectrometry, Anal Chem 2017;89:3829-3833.
68. Lima DB, de Lima TB, Balbuena TS et al. SIM-XL: A powerful and user-friendly tool for peptide cross-linking analysis, J Proteomics 2015;129:51-55.
69. Rost HL, Sachsenberg T, Aiche S et al. OpenMS: a flexible open-source software platform for mass spectrometry data analysis, Nat Methods 2016;13:741-748.
70. ME H. RNA Crosslinking Methods. In: ML (ed) Methods Enzymol.: Springer, 2009, 127-146.
71. Park AY, Robinson CV. Protein-nucleic acid complexes and the role of mass spectrometry in their structure determination, Critical Reviews in Biochemistry and Molecular Biology 2011;46:152-164.
72. Schmidt C, Kramer K, Urlaub H. Investigation of protein-RNA interactions by mass spectrometry--Techniques and applications, J Proteomics 2012;75:3478-3494.
73. Aoki T, Wolle D, Preger-Ben Noon E et al. Bi-functional cross-linking reagents efficiently capture protein-DNA complexes in Drosophila embryos, Fly (Austin) 2014;8:43-51.
74. Saliba AE, Vonkova I, Gavin AC. The systematic analysis of protein-lipid interactions comes of age, Nat Rev Mol Cell Biol 2015;16:753-761.
75. Gubbens J, Ruijter E, de Fays LE et al. Photocrosslinking and click chemistry enable the specific detection of proteins interacting with phospholipids at the membrane interface, Chem Biol 2009;16:3-14.
76. Yu SH, Bond MR, Whitman CM et al. Metabolic labeling of glycoconjugates with photocrosslinking sugars, Methods Enzymol 2010;478:541-562.
77. McCombs JE, Zou C, Parker RB et al. Enhanced Cross-Linking of Diazirine-Modified Sialylated Glycoproteins Enabled through Profiling of Sialidase Specificities, ACS Chem Biol 2016;11:185-192.
78. Bond MR, Zhang H, Vu PD et al. Photocrosslinking of glycoconjugates using metabolically incorporated diazirine-containing sugars, Nat Protoc 2009;4:1044-1063.
79. Toews J, Rogalski JC, Clark TJ et al. Mass spectrometric identification of formaldehyde-induced peptide modifications under in vivo protein cross-linking conditions, Anal Chim Acta 2008;618:168-183.
80. Arlt C, Gotze M, Ihling CH et al. Integrated Workflow for Structural Proteomics Studies Based on Cross-Linking/Mass Spectrometry with an MS/MS Cleavable Cross-Linker, Anal Chem 2016;88:7930-7937.
81. Mendes ML, Fischer L, Chen ZA et al. An integrated workflow for cross-linking/mass spectrometry, bioRxiv 2018:24.
Supplementary Table S1: Used vendor web page links.

| Vendor                  | Web site (accessed 02/2020)                                                                 |
|-------------------------|---------------------------------------------------------------------------------------------|
| Santa Cruz Biotechnology (SCB) | [http://www.scbt.com/chemicals-table-crosslinkers.html](http://www.scbt.com/chemicals-table-crosslinkers.html) |
| SCB                     | [http://www.scbio.de/chemicals-table-crosslinkers.html](http://www.scbio.de/chemicals-table-crosslinkers.html) |
| ChemCruz                | [http://www.scbio.de/chemicals-table-homobifunctional_crosslinkers.html](http://www.scbio.de/chemicals-table-homobifunctional_crosslinkers.html) |
| Invitrogen              | [https://www.thermofisher.com/search/global/searchAction.action?query=crosslinkers](https://www.thermofisher.com/search/global/searchAction.action?query=crosslinkers) |
| Sigma Aldrich           | [http://www.sigmaaldrich.com/life-science/molecular-biology/products.html?TablePage=9621740](http://www.sigmaaldrich.com/life-science/molecular-biology/products.html?TablePage=9621740) |
| Clear Synth             | [http://www.clearsynth.ca/search.asp?category=Cross%20Linking%20Reagents](http://www.clearsynth.ca/search.asp?category=Cross%20Linking%20Reagents) |
| ThermoFisher            | [https://www.thermofisher.com/search/results?query=crosslinkers&resultPage=1&resultsPerPage=60](https://www.thermofisher.com/search/results?query=crosslinkers&resultPage=1&resultsPerPage=60) |
| ThermoFisher            | [https://www.thermofisher.com/de/de/home/life-science/protein-biology/protein-labeling-crosslinking/protein-crosslinking/crosslinker-selection-guide.html](https://www.thermofisher.com/de/de/home/life-science/protein-biology/protein-labeling-crosslinking/protein-crosslinking/crosslinker-selection-guide.html) |
| ThermoFisher            | [https://www.thermofisher.com/de/de/home/life-science/protein-biology/protein-labeling-crosslinking/protein-crosslinking.html](https://www.thermofisher.com/de/de/home/life-science/protein-biology/protein-labeling-crosslinking/protein-crosslinking.html) |
| ThermoFisher            | [https://www.thermofisher.com/de/de/home/life-science/protein-biology/protein-labeling-crosslinking/protein-crosslinking/crosslinker-application-reference-guide.html](https://www.thermofisher.com/de/de/home/life-science/protein-biology/protein-labeling-crosslinking/protein-crosslinking/crosslinker-application-reference-guide.html) |
| ThermoFisher            | [https://www.thermofisher.com/de/de/home/life-science/protein-biology/protein-labeling-crosslinking/protein-labeling/biotinylation/photoreactive-biotinylation-reagents.html](https://www.thermofisher.com/de/de/home/life-science/protein-biology/protein-labeling-crosslinking/protein-labeling/biotinylation/photoreactive-biotinylation-reagents.html) |
| ThermoFisher            | [https://www.thermofisher.com/de/de/home/life-science/protein-biology/protein-labeling-crosslinking/protein-crosslinking/crosslinker-selection-tool.html](https://www.thermofisher.com/de/de/home/life-science/protein-biology/protein-labeling-crosslinking/protein-crosslinking/crosslinker-selection-tool.html) |
| ThermoFisher            | [https://www.thermofisher.com/us/en/products/99C8L0FC/crosslinking-labeling-protein-modification.html](https://www.thermofisher.com/us/en/products/99C8L0FC/crosslinking-labeling-protein-modification.html) |
| CovaChem                | [http://www.covachem.com/protein_crosslinkers.html](http://www.covachem.com/protein_crosslinkers.html) |
| CreativeMolecules       | [http://www.creativemolecules.com/CM_Products.htm](http://www.creativemolecules.com/CM_Products.htm) |
| ProteoChem              | [http://www.proteochem.com/proteincrosslinkers-c-1.html](http://www.proteochem.com/proteincrosslinkers-c-1.html) |
| G Biosciences           | [http://www.gbiosciences.com/ResearchProducts/Protein-Cross-Linking.aspx](http://www.gbiosciences.com/ResearchProducts/Protein-Cross-Linking.aspx) |
| G Biosciences           | [http://www.genotech.com/protein-research/protein-crosslinking-modification.html](http://www.genotech.com/protein-research/protein-crosslinking-modification.html) |
| G Biosciences           | [http://www.gbiosciences.com/Protein-Research/Cross-Linking-Modification/Protein-Cross-Linkers](http://www.gbiosciences.com/Protein-Research/Cross-Linking-Modification/Protein-Cross-Linkers) |
| Chemometec              | [http://shop.chemometec.com/?s=cross-linker&post_type=product](http://shop.chemometec.com/?s=cross-linker&post_type=product) |
| Interchim               | [http://www.interchim.eu](http://www.interchim.eu) |
| Biocompare              | [http://www.biocompare.com/Protein-Biochemistry/11119-Chemicals/?search=crosslinker](http://www.biocompare.com/Protein-Biochemistry/11119-Chemicals/?search=crosslinker) |
| BioVision               | [http://www.biovision.com/search/results.html?keywords=crosslinker&amp;x=0&amp;y=0](http://www.biovision.com/search/results.html?keywords=crosslinker&amp;x=0&amp;y=0) |
| Abcam                   | [http://www.abcam.com/products?keywords=crosslinker](http://www.abcam.com/products?keywords=crosslinker) |
| SoltecVentures          | [http://www.soltecventures.com/PhotoreactivePhenylAzides-c-8](http://www.soltecventures.com/PhotoreactivePhenylAzides-c-8) |
| MolBio                  | [http://www.molbio.com/HeteroBis](http://www.molbio.com/HeteroBis) |
| MolBio                  | [http://www.molbio.com/HomoNHS.htm](http://www.molbio.com/HomoNHS.htm) |
| Source       | URL                                                                 |
|--------------|----------------------------------------------------------------------|
| VWR          | https://de.vwr.com/store/product/18498564/cross-linkers-heterobifunctional |
| VWR          | http://www.molbio.com/Homomaleim.htm                                |
| VWR          | http://www.molbio.com/photoact.htm                                  |
| GooglePatents| https://www.google.de/?tbm=pts&gws_rd=cr&eil=CjIV6a7CYnWsAG2vYw#tbm=pts&q=cross-linker |
| GooglePatents| https://www.google.de/?tbm=pts&gws_rd=cr&ei=CjIV6a7CYnWsAG2vYw#tbm=pts&q=crosslinker |
| ChemicalBook | http://www.chemicalbook.com/Search_EN.aspx?keyword=crosslinker       |
| ChemicalBook | http://www.chemicalbook.com/ChemicalProductProperty_EN_CB9142806.htm  |
| ChemicalBook | http://www.chemicalbook.com/Search_EN.aspx                          |
| ChemCD       | http://www.chemcd.com/search?q=cross-linker                         |
| BioWorld     | https://www.bio-world.com/index.php?main_page=advanced_search_result&search_in_description=7&keyword=cross-linker&x=29&y=16 |
| VWR          | https://de.vwr.com/store/product/18498564/cross-linkers-heterobifunctional |
| CaymanChem   | https://www.caymanchem.com/Search?q=crosslinker                     |
| SoluLink     | http://www.solulink.com/search/google?cx=001277416618431018690%3Ahgffxulp8iy&cof=FORID%3A11%3BNB%3A1&query=crosslinker&op=Go&form_build_id=form-7d49f8e6f69c5b09c17b3c3c176a8bd0&form_token=fd2e01ee5d8c6941c678b5571f6972df&form_id=google_cse_searchbox_form |
| TRC (Toronto Research Chemicals) | http://www.trc-canada.com/products-listing/ |
| Vendor              | Publication, available on the web as PDF files                                                                 | web page (accessed 02/2020)                               |
|---------------------|-----------------------------------------------------------------------------------------------------------------|----------------------------------------------------------|
| Creative Molecules Inc. | Crosslinking reagents test reaction procedure                                                                   | http://www.creativemolecules.com                         |
| Creative Molecules Inc. | Isotopically-coded crosslinking reagents from Creative Molecules Inc., 2009                                      | http://www.creativemolecules.com                         |
| Creative Molecules Inc. | Crosslinking reagents masses                                                                                  | http://www.creativemolecules.com                         |
| GBiosciences        | Protein Cross Linker Selection Guide                                                                          | http://www.gbiosciences.com                              |
| GBiosciences        | Protein Cross-Linkers Handbook & Selection Guide                                                                | http://www.gbiosciences.com                              |
| GBiosciences        | Protein Cross-Linking Ttecher’s Guidebook (Cat. # BE-605)                                                      | http://www.gbiosciences.com                              |
| HiChrom             | Gas Chromatography GC Derivatization Reagents                                                                  | http://www.hichrom.com                                   |
| Invitrogen          | Molecular Probes Handbook, Chapter 5: Crosslinking and Photoactivatable Reagents, 11th. ed., 2010               | https://www.thermofisher.com                             |
| Interchim           | Derivatization reagents                                                                                       | http://www.interchim.com                                 |
| Mobitec             | Chapter 5: Crosslinking and Photoreactive Reagents                                                            | http://www.mobitec.com/cms/index.html                     |
| Pierce              | Cross-Linking Reagents, Technical Handbook, 2009                                                              | https://www.thermofisher.com                             |
| Regis               | GC Derivatization Reagents                                                                                     | http://www.registech.com                                 |
| Sigma Aldrich       | Derivatization reagents: For selective response and Detection in Complex Matrices                               | https://www.sigmaaldrich.com                             |
| Sigma Aldrich       | Guide to Derivatization Reagents for GC                                                                       | https://www.sigmaaldrich.com                             |
| Thermo Scientific   | Avidin-Biotin Technical Handbook                                                                             | https://www.thermofisher.com                             |
| Thermo Scientific   | Crosslinking Technical Handbook, 2012                                                                          | https://www.thermofisher.com                             |
| Thermo Scientific   | Thermo Scientific Pierce Reagents, Solvents and Accessories                                                  | https://www.cromlab.es                                   |
Supplementary Table S3: Used vendor-specific cross-linker selection tools resp. guides.

| Vendor          | Link to cross-linker selection tool / guide (accessed 02/2020) |
|-----------------|---------------------------------------------------------------|
| Thermo Scientific | [https://www.thermofisher.com/de/de/home/life-science/protein-biology/protein-labeling-crosslinking/protein-crosslinking/crosslinker-selection-tool.html](https://www.thermofisher.com/de/de/home/life-science/protein-biology/protein-labeling-crosslinking/protein-crosslinking/crosslinker-selection-tool.html) |
| Thermo Scientific | [https://www.thermofisher.com/de/de/home/life-science/protein-biology/protein-labeling-crosslinking/protein-crosslinking/crosslinker-selection-guide.html](https://www.thermofisher.com/de/de/home/life-science/protein-biology/protein-labeling-crosslinking/protein-crosslinking/crosslinker-selection-guide.html) |
| GBiosciences    | [http://info.gbiosciences.com/blog/bid/160707/What-is-Protein-Cross-Linking-and-Which-Reagents-are-Used-in-it](http://info.gbiosciences.com/blog/bid/160707/What-is-Protein-Cross-Linking-and-Which-Reagents-are-Used-in-it) |

Supplementary Table S4: Amino acid functional groups amenable to reactive groups of cross-linkers

| Amino acid | Functional group of amino acid ... is reacting primarily with the following reactive groups of cross-linkers |
|------------|--------------------------------------------------------------------------------------------------|
| S - serine | hydroxyl (-OH) isocyanate, vinyl sulfone                                                        |
| T - threonine | amine (-NH2), amides (formaldehyde, aryl halide, carbodiimide, carboxyl, hydroxybenzotriazole, phosphine, imidoester, N-hydroxysulfonyl, NHS ester, pentafluorophenyl, thioimidate, vinyl sulfone |
| Y - tyrosine | (formaldehyde, aryl halide, carbodiimide, carboxyl, hydroxybenzotriazole, phosphine, imidoester, N-hydroxysulfonyl, NHS ester, pentafluorophenyl, thioimidate, vinyl sulfone |
| K - lysine | amine (-NH2), amides (formaldehyde, aryl halide, carbodiimide, carboxyl, hydroxybenzotriazole, phosphine, imidoester, N-hydroxysulfonyl, NHS ester, pentafluorophenyl, thioimidate, vinyl sulfone |
| N - asparagine | amides (formaldehyde, aryl halide, carbodiimide, carboxyl, hydroxybenzotriazole, phosphine, imidoester, N-hydroxysulfonyl, NHS ester, pentafluorophenyl, thioimidate, vinyl sulfone |
| R - arginine | guanidino glyoxal, phenylglyoxal, formaldehyde                                                  |
| C - cysteine | sulfhydryl -SH acryl, haloacetyl (bromo-, iodo-), platinum(II), maleimide, methanethiosulfonate, pyridinylisulfide, vinyl sulfone |
| D – aspartic acid | carboxyl amine, carbodiimide, dimethylene                                                                 |
| E – glutamic acid | carboxyl amine, carbodiimide, dimethylene                                                                 |
| M - methionine | thioether (-S-) platinum(II)                                                                      |
| H - histidine | imidazole platinum(II), formaldehyde                                                               |
| W - tryptophan | indole formaldehyde                                                                               |
| F - phenylalanine | phenyl formaldehyde                                                                             |
Supplementary Table S5: Examples of typical spacer arm components, which are often used in cross-linking reagents

| Spacer arms          |
|----------------------|
| 6-amino-caprionic acid |
| DADPA (diaminodipropylamine) |
| ethylenediamine      |
| hexanediamine        |
| (PEG)<sub>n</sub> - polyethyleneglycol |

Supplementary Table S6: Handles for detection, affinity enrichment or purification used in cross-linking reagents

| Handle                  | Detection resp. enrichment / purification with ...                                                                 |
|-------------------------|----------------------------------------------------------------------------------------------------------------------|
| Biotin                  | avidin, streptavidin or neutravidin                                                                                   |
| alkyne                  | alkyne-azido click chemistry or immobilized azide linkers                                                             |
| azide (-N3)             | immobilized alkyne linkers or click chemistry using biotin-alkyne respective biotin-phosphine                           |
| CHCA (cyano-4-hydroxycinnamic acid) | UV-absorbing label for signal enhancement                                                                               |
| DNB (dinitrobenzole)    | anti-DNB antibodies                                                                                                   |
### Supplementary Table S7: Term statistics of the XLMOD.obo ontology

| Term type                                                                 | Number of terms |
|--------------------------------------------------------------------------|-----------------|
| Typedef's                                                                | 16              |
| Cross-linking attributes                                                 | 16              |
| Handles                                                                  | 10              |
| Reactive group terms                                                     | 39              |
| Reactive group attributes (without reactivities)                         | 8               |
| Reactivity terms                                                         | 37              |
| Cross-linker related chemical modification terms                         | 103             |
| Zero-length cross-linker                                                 | 6               |
| Bifunctional cross-linker terms                                          | 248             |
| Trifunctional cross-linker terms                                         | 13              |
| Label transfer reagents                                                  | 34              |
| Amino acid derivatives                                                   | 5               |
| Amino acids                                                              | 4               |
| GC/MS derivatization reagents                                            | 174             |
| LC/MS derivatization reagents                                            | 324             |
| other terms                                                              | 13              |
| Nucleobases and their derivatives                                        |                 |
| **Total number of terms**                                               | **1083**        |
| Term type                                                                 | ID range starting with |
|--------------------------------------------------------------------------|------------------------|
| General terms, reactivities, handles, other properties                    | xlmod:0000             |
| Concrete reactive groups and their attributes                            | xlmod:0010             |
| Mono-functional cross-linker (cross-linker related chemical modifications) | xlmod:0100             |
| Label transfer reagents                                                  | xlmod:0180             |
| Analogues of amino acids                                                | xlmod:0190             |
| Bi-functional cross-linkers                                              | xlmod:0200             |
| Tri-functional cross-linkers                                             | xlmod:0300             |
| Derivatization reagents                                                  | xlmod:0600             |
| Further reactivities (used for GC/MS and LC/MS reagents)                | xlmod:0650             |
| GC/MS derivatization reagents                                           | xlmod:0700             |
| LC/MS derivatization reagents                                           | xlmod:0900             |