Published in final edited form as:

Proteomics. 2012 February ; 12(3): 401–405. doi:10.1002/pmic.201100352.

An Enhanced Protein Crosslink Identification Strategy using CID-Cleavable Chemical Crosslinkers and LC/MSⁿ Analysis

Fan Liu^{1,†}, Cong Wu^{2,3,†}, Jonathan V. Sweedler², and Michael B. Goshe^{1,*}

¹Department of Molecular and Structural Biochemistry, North Carolina State University Raleigh, NC 27695-7622

²Institute for Genomic Biology, University of Illinois at Urbana-Champaign Urbana, IL 61801

³Department of Biochemistry, University of Illinois at Urbana-Champaign Urbana, IL 61801

Abstract

We describe a novel two-step LC/MSⁿ strategy to effectively and confidently identify numerous crosslinked peptides from complex mixtures. This method incorporates the use of our gas-phase cleavable crosslinking reagent, disuccinimidyl-succinamyl-aspartyl-proline (SuDP), and a new data processing algorithm CXLinS (Cleavable Crosslink Selection), which enables unequivocal crosslink peptide selection and identification on the basis of mass measurement accuracy, high resolving power, and the unique fragmentation pattern of each crosslinked peptide. We demonstrate our approach with well-characterized monomeric and multimeric protein systems with and without database searching restrictions where inter-peptide crosslink identification is increased 8-fold over our previously published data-dependent LC/MS³ method and discuss its applicability to other CID-cleavable crosslinkers and more complex protein systems.

Keywords

crosslinking; CID-cleavable crosslinker; proteomics; protein-protein interactions; mass spectrometry; mass measurement accuracy; multistage CID

Chemical crosslinking combined with mass spectrometry is a viable approach to characterize proteins, interactions and complex topologies [1-4]. Crosslinking reagents of various lengths and chemical functionalities react with the side chains of two distinct residues in close proximity within the same protein or between different protein molecules. The length of a given crosslinker provides a distance constraint which can be used to provide information on protein structure and interacting partners. However, due to the myriad of crosslinked products present at various stoichiometries, their identification by MS/MS analysis is difficult, particularly for inter-peptide crosslinks (which are the most structurally informative) due to simultaneously fragmenting both peptides, and thus preventing the individual peptide-product ion associations.

Considerable efforts have been directed to overcome these analytical challenges using two major approaches: (1) developing algorithms to facilitate interpretation of complex inter-peptide crosslink spectra and (2) designing novel crosslinking reagents to reduce sample complexity (e.g., incorporation of an affinity tag into the linker region) or to reduce data

* Address correspondence to: Dr. Michael B. Goshe Department of Molecular and Structural Biochemistry North Carolina State University 128 Polk Hall Campus Box 7622 Raleigh NC 27695-7622 phone: 919.513.7740 fax: 919.515.2047 michael_goshe@ncsu.edu.

[†]Each author contributed equally to this work

complexity (e.g., utilization of differential isotopically labeled or cleavable crosslinking reagents). One effective method has been the development of the program XQuest to identify inter-peptide crosslinks by combining the use of a novel search algorithm with isotope-coded crosslinkers [5]. However, the assignment of crosslinked peptides is entirely based on the algorithm, and the utilization of large databases is still a challenge due to the combinatorial increase of putative inter-peptide crosslinks. To enable further applications on native biological systems on the proteomic level, a crosslinking method that can efficiently identify inter-peptide crosslinks with high confidence is required.

To address these issues, our group has previously developed a gas-phase cleavable crosslinking reagent, disuccinimidyl-succinamyl-aspartyl-proline (SuDP), which incorporates a gas-phase labile bond aspartyl-prolyl amide bond (D-P) into the linker region [6]. This design renders the D-P bond in the crosslinked precursor (M_1 -SuDP- M_2) the preferential cleavage site during low energy CID, resulting in two highly abundant intact peptides with distinctive modifications (M_1 -SuD and P- M_2) to enable identification using intensity driven data-dependent acquisition (DDA) LC/MS³ analysis and database searching with Sequest [7, 8].

In this study, we developed a novel two-step LC/MSⁿ strategy to enhance inter-peptide crosslink identification with our CID-cleaved crosslinkers (Figure 1) using a hybrid 11 tesla LTQ-FT system (Thermo Fisher Scientific). For the first step, an LC/MS² experiment is performed under low-energy CID as described previously [6, 7] to favor the preferential cleavage of the D-P labile bond of the crosslinked species at the MS² stage, resulting in two distinctive peptides with signature fragmentation patterns. In our new approach, LC/MS² data are acquired with two essential features: (1) implementation of the “charge state rejection” function to preselect highly charged species (+4 and above) since most crosslinks are highly charged and (2) high mass measurement accuracy on both precursors and product ions to obtain accurate monoisotopic masses and charge state information via ProSightPC (Thermo Fisher Scientific) deconvolution. For the second step, a subsequent LC/MS³ analysis targeting each of the released peptides is performed based on the inclusion list generated from the LC/MS² data via our in-house developed algorithm CXLinkS (Cleavable Crosslink Selection). CXLinkS performs this operation by eliminating precursor ions identified by database searching with ProSightPC for uncrosslinked peptides and incorporating precursor ions whose product ions meet certain rules based on the chemistry of the CID-cleavable crosslinker (discussed below).

In CXLinkS a “pairing principle” is first applied for extraction of only product ion pairs whose mass sums are equal to their precursor ion masses (i.e., M_1 -SuDP- $M_2 = M_1$ -SuD + P- M_2 or complementary b- and y-ions). Next, an additional data screening is performed to further reduce the size of the inclusion list. Because SuDP is an asymmetric crosslinking reagent, a given lysyl residue can react with either the succinamyl (SuD) or prolyl (P) end of the molecule to generate two distinctive crosslinks (M_1 -SuDP- M_2 and M_2 -SuDP- M_1). The key feature of these two orientations is that they have exactly the same mass, and in most cases, the same LC retention time. Accordingly, an inter-peptide crosslink is defined as either “bi-directional”, if it provides four distinctive released peptides (M_1 -SuD, P- M_2 and M_2 -SuD, P- M_1), or “unidirectional” for two released peptides (M_1 -SuD, P- M_2 or M_2 -SuD, P- M_1). CXLinkS first searches for the “bi-directional” precursors based on the existence of a unique mass difference (i.e., $\Delta m = M_1$ -SuD – P- $M_1 = M_2$ -SuD – P- $M_2 =$ SuD – P) between two product ions. In this case, only four product ions (M_1 -SuD, P- M_2 , M_2 -SuD and P- M_1) from a specific precursor ion are selected for targeted LC/MS³ analysis. The remaining precursors are considered as potential “uni-directional” crosslinks and their released peptide selection is based on the intensity of the product ions. Due to the preferential cleavage of the D-P bond within the linker region under low-energy CID conditions, the two most intense

pairs are selected for targeted LC/MS³ analysis since these will represent potential released peptide pairs as opposed to b- and y-ion pairs. To setup the targeted data-dependent LC/MS³ experiment, the generated CXLinkS inclusion list containing *m/z* ratios, retention times, and charge states for selected MS precursor ions and *m/z* ratios for the MS² product ions is imported into the global inclusion list for the LC/MS³ data acquisition. For peptide identification, the MS³ product ion spectrum for each released peptide is interrogated by any suitable database searching algorithm.

We utilized our two-step LC/MSⁿ method on SuDP crosslinked BSA (a monomer) and human hemoglobin (an $\alpha_2\beta_2$ tetramer) using an 11 T LTQ-FT mass spectrometer. Each protein (10 μ M) was crosslinked with the SuDP reagent in PBS (pH 7.4) using a protein-to-crosslinker ratio of 1:100, then reduced, alkylated, and digested with trypsin (refer to Supplementary Methods for details). For LC/MSⁿ analysis, 2 pmol of an unfractionated and unenriched total peptide digest was used per injection. The MS² data were analyzed by ProSightPC and CXlinkS, and the MS³ spectra of BSA and human hemoglobin were searched against the Swiss-Prot *Bos taurus* and *Homo sapiens* database, respectively, using OMSSA (<http://pubchem.ncbi.nlm.nih.gov/omssa/>). To illustrate the efficacy of our method to identify inter-peptide crosslinks, data obtained from only one LC/MSⁿ analysis per sample were used (i.e., each sample consisting of one LC/MS² and one LC/MS³ analysis). In addition, various miscleavages of the same peptide sequence were counted as one unique crosslink and only those with the smallest number of miscleavages are reported. In total, 28 and 3 unique inter-peptide crosslinks were identified for crosslinked BSA (Supplementary Table 1) and hemoglobin (Supplementary Table 2) with no false positive identification from any other proteins. These identifications are highly confident because (1) each released peptide is identified individually based on its accurate mass (< 5 ppm) and corresponding fragmentation spectra and (2) to determine the identification of a crosslink, the released peptide pair selected by CXLinkS is required to be identified as one SuD-modified peptide and one P-modified peptide. Therefore, if the false positive rate (FPR) of identifying a peptide is *n* during traditional database searching, then *n*² is the FPR for the crosslink identification.

To further analyze the data, the MS³ spectra from these two crosslinking experiments were searched against a restricted database containing BSA or human hemoglobin. In addition to Lys residues, crosslinker modifications on His, Ser and Thr residues were applied and semi-tryptic cleavage (i.e., a tryptic cleavage site at either end of the peptide) was used. Also, the identification of only one of the SuD- or P-modified peptides via database searching was considered as originating from a potential inter-peptide crosslink, thus requiring the spectrum of its partner to be manually verified. Using this approach, 38 and 14 unique inter-peptide crosslinks were identified for BSA (Supplementary Table 3) and hemoglobin (Supplementary Table 4), respectively. For hemoglobin, this includes 6 unique inter-subunit crosslinks with an example of an identified bi-directional crosslink shown in Figure 2. All 6 inter-subunit crosslinks were mapped onto the structure of hemoglobin and are within the maximal distance constraint of 23.9 Å for the SuDP reagent (Figure 3). Compared to the results of searching against a large proteome database, the restricted database search provided an additional 10 and 11 identified crosslinks for BSA and hemoglobin, respectively. The difficulties of identifying both released peptides for these crosslinks from the MS³ product ion spectra from the unrestricted database searches were primarily due to (1) the short length of the peptides (less than 6 amino acids; 6 in BSA and 5 in hemoglobin), (2) the presence of semi-tryptically cleaved peptides (1 in BSA and 3 in hemoglobin), and (3) the peptides containing crosslink modifications on His, Ser, or Thr residues (3 in BSA and 5 in hemoglobin).

In our previous studies involving SuDP crosslinking of BSA and subsequent analysis using a DDA LC/MS³ approach [7, 8], we were only able to identify 5 inter-peptide crosslinks from BSA compared to the 38 using our new approach. This 8-fold increase is even more impressive when considering that the previous 5 crosslinks identified were obtained from four DDA LC/MS³ analyses using various dynamic exclusion times to promote selection of other ions from the MS² CID event as opposed to our new LC/MSⁿ approach requiring just one LC/MS² analysis and one LC/MS³ analysis. Unlike the DDA LC/MS³ analysis where the preponderance of crosslink identifications were dead-end products (i.e., a crosslink modification containing a hydrolyzed end) due to their high abundance compared to inter-peptide crosslinks [7, 8], our targeted LC/MS³ acquisition eliminates the triggering of MS³ CID events on dead-end products to promote fragmentation of the more structurally informative and less abundant inter-peptide crosslinks.

In conclusion, we developed a two-step LC/MSⁿ strategy using our previous designed CID-cleavable crosslinker SuDP. It significantly improves inter-peptide crosslink identification compared to our previous intensity-driven DDA approach where the MS² CID dissociation of the crosslinker is followed by selection of the top three most intense ions for MS³ analysis [7, 8]. More importantly, our LC/MSⁿ strategy and Δm principle to target released peptides for sequencing can be applied to other CID-cleavable crosslinkers [9, 10] using various high performance MSⁿ instruments such as LTQ-FT and LTQ-Orbitrap systems. Since there is no database size restriction for its application, the utilization of CID-cleavable crosslinkers and our two-step LC/MSⁿ strategy has the potential to be implemented for high-throughput proteomics studies to better understand protein-protein interaction networks.

Methods

Detailed methods are available in the Supplementary Methods section.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Tao Chen from the Laboratory of Immunology, NIAID, NIH for programming assistance as we developed the algorithm in CXLinks. We also thank the North Carolina Agricultural Research Service for continued support of our bioanalytical mass spectrometry research. Portions of this work were supported by a grant from the National Science Foundation (MCB-1021363) and National Institutes of Health (P30 DA018310).

References

1. 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. [PubMed: 16477643]
2. Jin Lee Y. Mass spectrometric analysis of cross-linking sites for the structure of proteins and protein complexes. *Mol Biosyst.* 2008; 4:816–823. [PubMed: 18633483]
3. Leitner A, Walzthoeni T, Kahraman A, Herzog F, et al. Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics. *Mol Cell Proteomics.* 2010; 9:1634–1649. [PubMed: 20360032]
4. Rappsilber J. The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. *J Struct Biol.* 2011; 173:530–540. [PubMed: 21029779]
5. Rinner O, Seebacher J, Walzthoeni T, Mueller LN, et al. Identification of cross-linked peptides from large sequence databases. *Nat Methods.* 2008; 5:315–318. [PubMed: 18327264]

6. Soderblom EJ, Goshe MB. Collision-induced dissociative chemical cross-linking reagents and methodology: Applications to protein structural characterization using tandem mass spectrometry analysis. *Anal Chem.* 2006; 78:8059–8068. [PubMed: 17134140]
7. Soderblom EJ, Bobay BG, Cavanagh J, Goshe MB. Tandem mass spectrometry acquisition approaches to enhance identification of protein-protein interactions using low-energy collision-induced dissociative chemical crosslinking reagents. *Rapid Commun Mass Spectrom.* 2007; 21:3395–3408. [PubMed: 17902198]
8. Liu F, Goshe MB. Combinatorial electrostatic collision-induced dissociative chemical cross-linking reagents for probing protein surface topology. *Anal Chem.* 2010; 82:6215–6223. [PubMed: 20560670]
9. Kao A, Chiu CL, Vellucci D, Yang Y, et al. Development of a novel cross-linking strategy for fast and accurate identification of cross-linked peptides of protein complexes. *Mol Cell Proteomics.* 2011; 10:M110 002212. [PubMed: 20736410]
10. Dreiocker F, Muller MQ, Sinz A, Schafer M. Collision-induced dissociative chemical cross-linking reagent for protein structure characterization: applied Edman chemistry in the gas phase. *J Mass Spectrom.* 2010; 45:178–189. [PubMed: 19950134]

Figure 1.

Workflow for the identification of inter-peptide crosslinks using our LC/MSⁿ approach. For LC/MS² analysis, precursors that are selected for MS² fragmentation include two types: crosslinked peptides (M₁-SuDP-M₂ as red and M₂-SuDP-M₁ as green) and uncrosslinked peptides (M₃ as grey). The uncrosslinked peptides are identified by ProSightPC. The crosslinked peptides are further analyzed by CXLinkS. CXLinkS divides inter-peptide crosslinks into two groups, “bi-directional” and “uni-directional” according to the orientation of the crosslinker. The bi-directional form contains two orientations of the crosslink (M₁-SuDP-M₂ as red and M₂-SuDP-M₁ as green), which is dissociated into four major components (M₁-SuD, P-M₂, M₂-SuD and P-M₁) during MS² fragmentation (low

energy CID). The uni-directional form contains one orientation of the crosslink (M_1 -SuDP- M_2 as red), which is dissociated into two components (M_1 -SuD and P- M_2) during MS² fragmentation (low energy CID). All the selected components (red, green, and blue) are targeted for MS³ analysis (high energy CID) in the subsequent LC/MS³ experiment based on the inclusion list generated by CXLinkS to enable crosslinked peptide identification using database searching of the MS³ spectra (red and green).

Figure 2.

LC/MSⁿ identification of a bi-directional SuDP inter-subunit crosslink in human hemoglobin. The (A) MS spectrum shows the intact +5 inter-peptide crosslink precursor $[M_1\text{-SuDP-M}_2\text{+5H}]^{5+}$ and $[M_2\text{-SuDP-M}_1\text{+5H}]^{5+}$. On the basis of the charge state and exclusion time restriction, this precursor was selected as the most intense eligible ion and fragmented during the MS² low energy CID event primarily at the D-P bond to produce four distinct peptides ($[M_1\text{-SuD+2H}]^{2+}$, $[P\text{-M}_2\text{+3H}]^{3+}$, $[P\text{-M}_1\text{+2H}]^{2+}$, and $[M_2\text{-SuD+2H}]^{2+}$) as the predominant ions in the (B) MS² spectrum. These four ions were selected by CXLinkS and used for the targeted LC/MS³ identification to produce the (C) MS³ spectra (annotated MS³ spectra appear in Supplementary Figure 1). By virtue of the unique mass produced by collision-induced dissociation of the crosslinker (P: 97.0528 u; SuD: 197.0324 u), OMSSA (<http://pubchem.ncbi.nlm.nih.gov/omssa/>) was used to identify the peptide sequence and the site of modification.

Figure 3.

Mapping of identified SuDP inter-subunit crosslinks of human hemoglobin. The inter-subunit crosslinks identified by our LC/MSⁿ approach were mapped onto the 1.74 Å resolution X-ray crystallographic structure of human deoxyhemoglobin (PDB code: 2HHB). Hemoglobin is a heterotetramer comprised of two identical α subunits (cyan) and two identical β subunits (blue). All unique identified crosslinked residues are indicated in red and the corresponding residues due to structural symmetry are indicated in gray. Unique (A) α - β crosslinks, (B) β - β crosslinks and (C) α - α crosslinks are indicated with light yellow lines and those due to structural symmetry are indicated with dark yellow lines. Distances were measured between the two α -carbons at the corresponding residues to ensure each crosslink was within the maximum distance constraint of 23.9 Å for the SuDP reagent. The images were generated using PyMOL version 1.2.8 (<http://www.pymol.org/>).