1 **Performing Protein Crosslinking using Gas-phase Cleavable Chemical**
2 **Crosslinkers and Liquid Chromatography-Tandem Mass Spectrometry**

3
4
5 Andrew S. Argo, Chunxiao Shi, Fan Liu, and Michael B. Goshe*

6
7
8 Department of Molecular and Structural Biochemistry
9 North Carolina State University
10 Raleigh NC 27695-7622

11
12
13
14
15
16
17 *Address correspondence to:
18 Dr. Michael B. Goshe
19 Department of Molecular and Structural Biochemistry
20 North Carolina State University
21 128 Polk Hall
22 Campus Box 7622
23 Raleigh NC 27695-7622
24 phone: 919.513.7740
25 fax: 919.515.2047
26 email: michael_goshe@ncsu.edu
27
28

Abstract

In this article, we describe our methods and protocols using collision-induced dissociative chemical crosslinking-tandem mass spectrometry (CID-CXL-MS/MS) analysis and the practical considerations when implementing these reagents and methodology for protein crosslinking studies. The synthesis of our novel chemical crosslinkers is described as well as their use for effectively labeling protein and protein complexes. Several sample preparation methods for liquid chromatography-tandem mass spectrometry are provided including the enrichment of interpeptide crosslinks. For identification of CID-CXL-MS/MS crosslinks, details regarding MS acquisition parameters and the utilization of various mass spectrometers are addressed along with post-data acquisition analysis to identify interpeptide crosslinks. Once the CID-CXL-MS/MS approach is optimized for a protein target or a set of targets, it can be used as a tool for biological research for studying protein structure when integrated with data obtained using other techniques, such as NMR and X-ray crystallography, or extended to the study of protein-protein interactions in physiological environments.

Keywords

Protein structure, protein-protein interactions, crosslinking, gas-phase cleavable crosslinkers, CID-cleavable crosslinkers, mass spectrometry

1. Introduction

Chemical crosslinking (CXL) combined with mass spectrometry (MS) is a viable approach to study the low-resolution structure of proteins and protein complexes that provides insight regarding protein interactions and topology [1-3]. These measurements require relatively

small amounts of sample and can be performed under a wide range of conditions that mimic native protein environments. CXL-MS becomes an even more powerful tool in structural biochemistry when combined with other techniques such as NMR, X-ray crystallography, and cryo-electron microscopy ([4-6]. Since CXL is not limited by the size of proteins or protein complexes, it allows for a large variety of proteins and protein systems to be studied. For example, membrane proteins have posed challenges to X-ray crystallography and NMR structural analysis due to their solution instability and large size compared to cytosolic proteins. As a result, structural analysis of membrane proteins is often performed in a systematic fashion by cloning the specific domains of these proteins to obtain structures; however, both structural and functional information for the intact protein is lost using this divide-and-conquer approach. However, by applying CXL-MS on the intact protein in its native environment in combination with the experimentally determined domain structures, spatial domain orientation can be performed using molecular modeling to construct a working structural model that can be implemented to design biochemical experiments to study protein function.

CXL-MS is composed of two distinct experiments: (1) chemical crosslinking of the proteins and (2) identification of the crosslinked peptide products using liquid chromatography-tandem mass spectrometry (LC/MS/MS). CXL experiments usually involve the use of homo- and hetero-bifunctional crosslinking reagents to covalently attach the side chains of two protein residues within a defined spatial proximity [1, 7]. Based on the fully extended length of the side chains between two α -carbons and the length of the corresponding crosslinker, a maximum distance constraint can be determined for the two crosslinked residues which provides information on the spatial orientation of protein structural elements and their interactions. The key to a successful protein crosslinking experiment is to provide conditions that allow the protein

1 to exist in its physiological environment while promoting crosslinker chemical reactivity that
2 preserves the overall tertiary structural integrity of the protein. Conditions that over-crosslink a
3 protein or protein complex will produce crosslinked products that do not accurately reflect native
4 structure due to label-induced denaturation, disassembly, or aggregation. Consequently,
5 conditions to properly crosslink proteins for a given crosslinker need to be considered in order to
6 perform a successful crosslinking experiment.

7 Another issue associated with performing a successful CXL-MS experiment is related to
8 the myriad of crosslinked modifications that are produced on proteins subjected to CXL. After a
9 crosslinking reaction is performed, the proteins are proteolytically digested (either in-gel or in-
10 solution) and the generated peptide products are then analyzed by reversed-phase LC/MS/MS.
11 The proteolytic digestion of crosslinked proteins produces a complex mixture containing highly
12 abundant unmodified peptides and lower abundant crosslinked peptide products which are
13 primarily comprised of three types (Figure 1). The most abundant crosslinked products are the
14 deadend crosslinks which are chemically the result of one end of the crosslinker reacting with an
15 amino acid residue of the protein and the other end remaining unreacted or hydrolyzed with
16 water (or some other quencher) as opposed to reacting with another protein amino acid residue.
17 The crosslinked products of low stoichiometry are intrapeptide crosslinks, which contain a
18 crosslink between two residues on the same peptide, and interpeptide crosslinks (the most
19 structurally informative), which contain a crosslink between two peptides that can be formed on
20 the same protein or between two different proteins. Depending on the protease used for the
21 digestion, a different number of intrapeptide and interpeptide crosslinks can be generated. Once
22 identified, these crosslinks are used to define a set of distance constraints which can then be
23 applied to probe protein structural features.

Although CXL-MS has been a useful approach to study protein structure, the identification of crosslinked products remains an analytical challenge for several reasons. During LC/MS/MS analysis, the products of the proteolytic digest are detected as precursor ions and their corresponding product ions produced by collision-induced dissociation (CID), or some other mechanism of fragmentation, are used to identify the unmodified peptides and crosslinked products. The complexity associated by the introduction of a chemical crosslink makes unequivocal identification difficult due to charge competition during electrospray ionization and the low stoichiometry of crosslinked products compared to the preponderance of theoretical products. In addition, the interpretation of product ion spectra of interpeptide crosslinks is more complex compared to modified peptides, such as deadend and intrapeptide crosslinks, due to the increased number of product ions that can be generated. To enable confident interpretation of complex product ion spectra of interpeptide crosslinks, the use of mass spectrometers capable of high mass measurement accuracy and high resolving power is required. Database searching algorithms such as SEQUEST and Mascot cannot interpret interpeptide crosslink product ion spectra, and thus a considerable amount of effort has been directed towards developing algorithms to facilitate peptide crosslink identification such as GPMAW [8], MS2Assign[9], SearchXLinks [10], VIRTUALMSLAB [11], xQuest[12], DBond [13], Crux Search-for-Xlinks[14], pLink[15], xProphet[16], StavroX[17], DXMSMS Match [18, 19], MeroX[20], SIM-XL[21].

In addition to improvements in database searching algorithms to identify protein crosslinks by LC/MS/MS, gas-phase cleavable crosslinkers have been developed to promote identification and have many features in common [22-28]. A decade ago we developed a low energy collision-induced dissociative chemical crosslinking reagents, disuccinimidyl-

1 succinamyl-aspartyl-proline (SuDP) and disuccinimidyl-succinamyl-aspartyl-prolyl-glycine
2 (SuDPG) (Soderblom, 2006 #10} and a multistage activation approach to permit unequivocal
3 identification of deadend, intrapeptide and interpeptide crosslinks using an LCQ Deca ion trap
4 mass spectrometer [23]. Since then, we have enhanced interpeptide detection using hybrid ion
5 trap-ICR and ion trap-orbitrap instruments [29]. These reagents have an aspartyl-prolyl bond
6 within the linker region which can be selectively dissociated under low-energy conditions at the
7 MS² stage to produce two unique modified peptides that are further subjected to CID during a
8 subsequent MS³ event. Database searching of the generated MS² and MS³ product ion spectra
9 enables the identification of deadend/intrapeptide and interpeptide crosslinks, respectively [23].
10 We also synthesized another crosslinker, disuccinimidyl-succinamyl-valyl-proline (SuVP) to
11 complement the SuDP reagent and used both as a combinatorial pair to probe the topological
12 electrostatic surface potential of proteins [27] and have applied our CXL approach to probe the
13 domain orientation of AbrB to assist in NMR structure determination of the tetramer [4].

14 In this presentation, we describe our collision-induced dissociative chemical crosslinking-
15 tandem mass spectrometry (CID-CXL-MS/MS) approach to promote better identification of
16 crosslink modifications [4, 22, 23, 27, 29]. First, we will present how we establish conditions for
17 chemical crosslinking and sample preparation for LC/MS/MS analysis. Second, we describe the
18 LC/MS/MS data acquisition methods for interpeptide crosslink analysis. Third, we will address
19 the post-acquisition data analysis using the CID-CXL-MS/MS reagents. Although the results
20 obtained using our CID-CXL-MS/MS approach can be directly applied for protein structural
21 analysis, our reagents and its design concept can potentially be used to develop a more elaborate
22 and comprehensive analysis of protein interactions and networks in order to quantify protein
23 conformational dynamics and the kinetic changes occurring in a cellular interactome.

2. Materials

2.1 Chemicals and reagents

Bovine serum albumin (BSA) 98% (fatty acid- and globulin-free), dimethyl sulfoxide (DMSO), formic acid (ACS reagent grade), dithiothreitol (DTT), iodoacetamide, acetonitrile (HPLC grade), N-hydroxybenzotriazole, diisopropylcarbodiimide, and trifluoroacetic acid (TFA) were from Sigma-Aldrich (www.sigmaaldrich.com). Glycine, sodium chloride, ethylenediaminetetraacetic acid (EDTA), and disuccinimidylsuberate (DSS) were from Thermo Fisher Scientific (www.thermofisher.com). Di-sodium hydrogen phosphate dehydrate, sodium dihydrogen phosphate monohydrate, ammonium bicarbonate were from Fluka (www.sigmaaldrich.com). Tris-HCl and sequencing-grade modified trypsin were from Promega Life Sciences (www.promega.com). Urea (99%) was from Acros Organics (www.acros.com). Vivacon 500 10kDa MWCO HY filters were from Sartorius Stedim Biotech (www.sartorius.com). Fmoc-Pro-NovaSyn TGT extremely acid-sensitive preloaded resin, Fmoc-Gly-NovaSyn TGT extremely acid sensitive preloaded resin, Fmoc-Pro-OH, Fmoc-Asp(OtBu)-OH, and Fmoc-Val-OH were from EMD Biosciences (www.emdbiosciences.com). Water was distilled and purified using a High-Q 103S water purification system (www.high-q.com). All other reagents and chemicals were from Sigma-Aldrich-Fluka or Thermo Fisher Scientific and used with further purification unless otherwise noted.

2.2 Synthesis of CID-CXL-MS/MS crosslinkers

The SuDP, SuDPG, and SuVP reagents are synthesized and purified as described previously [22, 23, 27]. Synthesis of SuDP is performed by first coupling the Fmoc-Asp(OtBu)-OH to the Fmoc-Pro-NovaSyn TGT extremely acid-sensitive preloaded resin in a batch wise manner using the active ester method. Following deprotection of the resin-bound amino acid, active esters are formed using Fmoc-amino acid, N-hydroxybenzotriazole, and diisopropylcarbodiimide in a 5:5:5.5 molar ratio to the resin-bound amino acid. Following Fmoc deprotection of Fmoc-Asp(OtBu)-OH, succinic anhydride is added to introduce a carboxylate group to the N-terminus of the reagent linker region. The completed linker region is cleaved from the resin with 1% TFA in dichloromethane. Following reversed-phase high-performance liquid chromatography (HPLC) purification, esterification of the dicarboxylic acid is performed using either di(N-succinimidyl) oxalate or di(N-succinimidyl) carbonate. The esterified product is purified via HPLC, dried under vacuum, and stored at -20 °C. Before use, the aspartyl side chain tert-butyl protecting group is removed by the addition of 90% TFA/5% water/5% DMSO (v/v/v), and the mixture is stirred at ambient temperature for 10 min, then dried under vacuum. The SuDP crosslinker is purified via HPLC to remove deprotection contaminants, dried under vacuum, and then is stored at -80°C.

Synthesis of SuVP is performed in the same manner as the SuDP reagent using Fmoc-Val-OH in place of Fmoc-Asp(OtBu)-OH; however the final purification of the SuVP reagent is performed via HPLC immediately following esterification. Synthesis of SuDPG is performed in a similar fashion as SuDP except using Fmoc-Gly-NovaSyn TGT extremely acid sensitive preloaded resin and Fmoc-Pro-OH to introduce the additional glycyl moiety.

Quantification of the SuDPG, SuDP, and SuVP reagents is performed by measuring the absorbance obtained at 260 nm of a small aliquot from the HPLC-purified product and

calculating its concentration based on a standard curve generated from various concentrations of disuccinimidyl suberate (DSS) and saponification of the NHS-esters using NaOH [30]. The structures of all three reagents and their extended crosslinker lengths used for calculating maximum distance constraints are presented in Figure 2.

3. Methods

The methods described below pertain to those previously developed to study protein structure [4, 22, 23, 27, 29] and those currently being utilized in our laboratory. The general workflow of the CID-CXL-MS/MS approach (Figure 3) consists of consecutive steps of protein crosslinking, possible purification of the crosslinked proteins, enzymatic digestion, separation and enrichment of the crosslinked peptides, followed by reversed-phase liquid chromatography-tandem mass spectrometry analysis using LC/MS² and LC/MS³ acquisition methods. The protocols for these steps are described as well as alternatives pertaining to the CID-CXL-MS/MS approach. It should be noted that before attempting an analysis of a protein or protein system of interest, the protocols chosen for a particular CID-CXL-MS/MS experiment should be evaluated using bovine serum albumin (BSA) as a control.

3.1. Protein crosslinking with CID-CXL-MS/MS crosslinkers

A typical protein crosslinking experiment with our CID-CXL-MS/MS reagents [4, 22, 23, 27, 29] usually contains 10 μ M of protein in 100 mM sodium phosphate, 25 mM NaCl, 1.0 mM EDTA, pH 7.5. Aliquots of our CID-CXL-MS/MS reagent are dissolved in DMSO and added to the protein solutions to produce a protein-to-crosslinker molar ratio (defined as the ratio of the number of moles of protein to the number of moles of crosslinker) of 1:0 (control), 1:50, 1:100,

1 1:200, and 1:300 while keeping the final organic concentration below 1-3% of the total reaction
2 volume. The crosslinking reaction is performed at ambient temperature for 30 to 60 min
3 followed by quenching with 1.0 M Tris or 1.0 M glycine, pH 7.5, to produce a final
4 concentration of 10-50 mM, and then the quenching reaction is incubated for 15-30 min at
5 ambient temperature. Samples are then analyzed by SDS-PAGE using a 4-20% Bis-Tris gel
6 (Invitrogen Life Technologies, www.lifetechnologies.com), and proteins are visualized with
7 Coomassie staining. Based on this analysis, appropriate protein-to-crosslinker molar ratios
8 (typically 1:100 or 1:200 for proteins of the size of BSA, but these ratios will vary depending on
9 the size and number of reactive amino acid residues of the proteins being analyzed) will be
10 selected for either in-gel, in-solution or filter-assisted proteolytic digestion.

12 ***3.2 Sample preparation of proteins crosslinked with CID-CXL-MS/MS crosslinkers***

13 Depending on the nature of the protein sample after crosslinking, samples can be
14 prepared using in-gel, in-solution, or filter-aided protease digestion as described below.

15 Although these methods use trypsin digestion, other specific and nonspecific proteases can be
16 used to provide an alternative for identify additional crosslinked products [22, 31].

18 ***3.2.1. In-gel sample preparation***

19 To assess the extent of crosslinking and to separate crosslinked complexes from
20 uncrosslinked components, SDS-PAGE is very effective [4, 22, 23]. After quenching the CID-
21 CXL-MS/MS crosslinking reaction with glycine or Tris-HCl, the solvent is removed via vacuum
22 centrifugation to near dryness (about 10 µl remaining), which is then diluted with SDS PAGE
23 running buffer. Depending on the complexity of the protein system and the volume of the CID-

CXL-MS/MS crosslinking reaction, the sample is diluted with the appropriate amount of loading buffer and an aliquot sufficient to visualize the proteins by Coomassie staining is then loaded for each lane of a 15% SDS PAGE gel. After electrophoresis, protein bands are visualized by Coomassie staining.

To identify crosslinked peptide products for the separated protein complexes, in-gel digestion can be performed using a modified protocol previously described [32]. Protein bands of interest are excised using a razor blade or X-Acto knife and are destained using 1 ml of 40% acetonitrile/60% 50 mM ammonium bicarbonate, pH 8.0. After 1 h of destaining with gentle mixing using a rocker, each gel slice is reduced for 1 h at 37°C with 5 mM DTT in 50 mM ammonium bicarbonate, pH 8.0, and then cysteinyl residues are alkylated with 20 mM iodoacetamide in 50 mM ammonium bicarbonate, pH 8.0, at ambient temperature in the dark for 1 h. In-gel digestion is performed overnight at 37°C using trypsin resuspended in 50 mM ammonium bicarbonate, pH 8.0, at a final 1:5 (w/w) enzyme-to-protein ratio. Peptides are extracted from the gel pieces three times with 200 µl of 60% acetonitrile/35% 50 mM ammonium bicarbonate, pH 8.0/5% formic acid (v/v/v) with sonication in a water bath for 20 min. Extractions are pooled, dried under vacuum centrifugation, and stored at -80°C.

3.2.2. In-solution sample preparation

After quenching the CID-CXL-MS/MS crosslinking reaction with glycine or Tris-HCl, the sample can be subjected to in-solution proteolytic digestion as previously described [4, 22, 23, 27, 29]. A typical protocol is described here, but it can be modified as needed, depending on the extent of chemical denaturation, disulfide bond reduction and cysteinyl alkylation that is required for the protein/protein complex being studied. The samples are reduced with 40 mM

DTT for 1 h at 37°C and cysteinyl residues are alkylated in the dark for 1 h at ambient temperature with 200 mM iodoacetamide. To remove excess reagents, the sample is dialyzed overnight against 50 mM ammonium bicarbonate, pH 8.0, using a 3000 MWCO Slide-A-Lyzer dialysis cassette (Pierce Biotechnology (www.piercenet.com)). Dialyzed samples are digested in solution overnight at 37°C with trypsin using an enzyme-to-protein ratio of 1:50 (w/w) then dried via vacuum centrifugation and stored at -80°C. To provide better protease digestion of crosslinked proteins, the sample can first be denatured by adding solid guanidine-hydrochloride (Gdn-HCl) to make a concentration of 9 M Gdn-HCl following by incubating the sample in a boiling water bath for 5 min; then the reduction, alkylation, dialysis, and proteolytic digestion are performed as described above. Alternatively, after reduction and alkylation, the sample can be diluted to 1 M Gdn-HCl by adding 50 mM ammonium bicarbonate, pH 8.0, prior to tryptic digestion; following digestion, the peptides can be purified by solid-phase extraction (SPE) using a Grace Alltech Prevail C₁₈ Extract-Clean column (www.grace.com) connected to a vacuum manifold, dried via vacuum centrifugation, and stored at -80°C.

3.2.3. Filter-aided sample preparation (FASP)

To facilitate a more efficient sample workup of multiple crosslinked samples, we have been using filter-aided sample preparation (FASP) as previously described for proteomic analysis [33]. After quenching the CID-CXL-MS/MS crosslinking reaction with glycine or Tris-HCl, the sample is reduced with 100 mM dithiothreitol to reach a final concentration of 5 mM and incubated at 56°C for 30 min. After incubation, the crosslinked sample is mixed with 200 µl of 8 M urea and transferred to a Vivacon 500 10kDa MWCO filter column (a smaller or larger MWCO filter can be used, depending the nature of the protein sample). The filtering device is

centrifuged at ambient temperature for 30 min at 14,000xg. Then the retained protein on the filter is washed with 200 μ l of 8 M urea and the flow-through is discarded. For alkylation, 100 μ l of 50 mM iodoacetamide is added, and the whole filter unit is placed in the dark for 20 min at ambient temperature followed by centrifugation at 14,000xg for 20 min. The concentrate is diluted with 100 μ l of 8 M urea and centrifuged for another 20 min and this step is repeated two additional times. Afterwards, 100 μ l of 50 mM ammonium bicarbonate, pH 8.0, is added and centrifuged for another 20 min and this step is repeated two additional times. The collection tube used for the washes is replaced with a new tube, then 10 μ l of trypsin solution (1.0 μ g/ μ l in resuspension buffer provided by the manufacturer) and 40 μ l of 50 mM ammonium bicarbonate, pH 8.0, are mixed gently and placed on the filter, then the filter-and-tube assembly is placed in a float and incubated for 4-18 h in a water bath at 37 °C. The digested protein sample is then removed from the filter with 80 μ l of 50 mM ammonium bicarbonate, pH 8.0. After centrifugation, the solvent is evaporated (or reduced in volume) via vacuum centrifugation and stored at -80°C.

3.3. Strong cation exchange enrichment of interpeptide crosslinks

Although CID-CXL-MS/MS crosslinked samples that have been proteolytically digested can be analyzed directly by LC/MS/MS, the sample complexity introduced by crosslinking and the low stoichiometry of interpeptide crosslinked products may require the implementation of strong cation ion exchange (SCX) to separate low charged peptides from higher charged species, thus enriching for interpeptide crosslinked products[12]. Peptides from proteolytic digests (these samples may have to be purified by SPE if they contain any competing cations such as Tris or guanidium) are separated by SCX using a pipette tip in-house packed with Poros XS material (www.appliedbiosystems.com). The peptide mixture is loaded onto the SCX column, and then

bound peptides are step-wise eluted with 100 mM, 200 mM, 300 mM and 1 M NaCl containing 0.1% TFA. Peptides from each fraction were desalted using a separate SPE pipette tip in-house packed with Poros OLIGO R3 material (www.appliedbiosystems.com), dried via vacuum centrifugation, and stored at -80°C.

3.4 Liquid chromatography-tandem mass spectrometry analysis of CID-CXL-MS/MS crosslinked samples

Reversed-phase separation of peptides obtained from CID-CXL-MS/MS crosslinked protein proteolytic digests can be performed on any suitable liquid chromatograph coupled to a tandem mass spectrometer [4, 22, 23, 27, 29]. A typical gradient that we have used for analyzing crosslinked samples implementing in-house packed C₁₈ capillary columns is as follows. The mobile phases consist of (A) 0.1% formic acid in water and (B) 0.1% formic acid in acetonitrile. After loading approximately 1.0 nmol of total peptide digest onto the reversed-phase column (e.g., 150 µm i.d. x 40 cm length fused-silica capillary slurry-packed with 5 µm 300 Å Jupiter C18 stationary phase (Phenomenex, www.phenomenex.com)) for microLC separations (1-5 µl/min flow rates) or 200 pmol onto the reversed-phase column (e.g., 75 µm i.d. x 25 cm length fused-silica capillary slurry-packed with 3 µm 200 Å Magic C18AQ stationary phase (Michrom Bioresources Inc., <http://www.michrom.com>)) for nanoLC separations (100-300 nl/min flow rates), peptides are eluted using a linear gradient of 1.0% B/min to a high percentage of B (up to a range between 40 and 95% B) and electrosprayed in the positive ion mode. For our nanoLC separations, a trapping column (e.g., 100 µm i.d. x 2 cm length Acclaim PepMap 100 trap column packed with 5 µm 100 Å C18 stationary phase (Thermo Scientific, www.thermoscientific.com) is typically coupled to the analytical column. Depending on the

operation of the LC system, the mobile phases, gradient, flow rate, and electrospray voltage will need to be established in order to provide an appropriate reversed-phase separation for LC/MS/MS analysis.

3.5 Mass spectrometry data acquisition

Although reversed-phase separation of CID-CXL-MS/MS crosslinked peptides is an important aspect of LC/MS/MS analysis, the MS data acquisition is even more important in order to take advantage of the gas-phase cleavable crosslinkers for identifying protein crosslinks and involve either an LC/MS² or LC/MS³ data-dependent acquisition.

3.5.1 Data-dependent acquisition using LC/MS²

For the LC/MS² acquisitions, both MS and MS² events need to be performed on tandem mass spectrometers capable of high resolving power with high mass measurement accuracy. Based on the acquisition settings using an ion trap-ICR instrument [29], acquisitions using an ion trap-orbitrap (Thermo Scientific Orbitrap Elite) were established and are presented here, but they can be modified as required for other systems. Both MS and MS² scan events are performed in the orbitrap. MS scans (m/z 400-2000) are acquired after accumulation to an automatic gain control (AGC) target value of 1×10^6 in the linear ion trap. The three most abundant ions with charge states ≥ 4 are sequentially selected using an isolation width of m/z 7 with an AGC target value of 1×10^5 and fragmented by collision-induced dissociation (CID) using a normalized collision energy setting of 35% or high-energy collision-induced dissociation (HCD) using a normalized collision energy setting of 25%. It should be noted that selecting ions for CID or HCD with charge states ≥ 4 precludes the trapping of intrapeptide and deadend crosslinker

modified precursor ions that often have charge states ≤ 3 . Typically, a minimum signal threshold of 1×10^4 is used for MS^2 activation, but can be increased to 2×10^4 in an effort to generate product ions of higher intensity. Dynamic exclusion settings will depend upon the sample complexity and the resolution of the separation achieved during the LC gradient. For some of our acquisitions, we have used a dynamic exclusion repeat count of 2, a dynamic exclusion repeat duration of 60 s, and an exclusion duration of 90 s. The maximum allowed ion accumulation times of 200 ms for the MS scans and 500 ms for the MS^2 scans are employed with the resolving power set to 60,000 for the MS scans and 15,000 for the MS^2 scans at m/z 400.

3.5.2 Data-dependent acquisition using LC/ MS^3

Ion trap CID using data-dependent multiactivation (LC/MS^3) was performed as previously described [23, 27]. A series of continuously repeating scan events is employed to promote CID-CXL-MS/MS crosslink product detection. A survey MS scan is performed from m/z 400 to 2000 for the first scan event. The second scan event (the CID dissociation scan for the precursor crosslinked species to promote crosslinker cleavage), the most abundant ion from the survey MS scan is data-dependently selected and subjected to CID for MS^2 analysis. For the next series of scan events (at least two, but more may be used for systems capable of microsecond scanning), the most abundant ions from the second scan are then data-dependently selected and individually subjected to CID for MS^3 analysis. All CID events can be performed using the typical settings employed for linear peptide analysis (e.g., a normalized collision energy setting of 35% on Thermo Scientific ion traps), however a lower value of 25% (which represents around a 30% reduction for this instrument setting) implemented on scan event two (i.e., the crosslinker dissociation event) can be used to minimize b- and y-ion formation to

promote the detection of the modified linear peptides generated from the CID-CXL-MS/MS interpeptide crosslink during the subsequent MS³ scan events, as shown previously [22, 23]. For MS² and MS³ activation on most modern linear ion traps the minimum signal threshold of 1×10^4 and 1×10^3 , respectively, is typically used. To increase the detection and identification of lower abundant interpeptide crosslinked precursor ions, data are acquired using dynamic exclusion (as indicted in Section 3.5.1, dynamic exclusion settings can vary based on sample complexity and the resolution of the LC separation). For example, separate acquisitions can be performed using no dynamic m/z exclusion list and a 12 s dynamic m/z exclusion list[27] but may be altered as needed due to sample complexity.

3.6 Post-acquisition data analysis

3.6.1 Analysis of crosslinked products using data obtained from LC/MS² acquisitions

Data collected from LC/MS² experiments of crosslinked protein samples labeled with cleavable and non-cleavable crosslinkers can be analyzed by a variety of software packages that are publically available on the web and those listed in the Introduction (Section 1) and continue to grow in number. One we have used successfully with our CID-CXL-MS/MS approach for the analysis of interpeptide crosslinks is MeroX [20] which is specifically designed for cleavable crosslinkers. For analysis, acquired LC/MS² data are processed to generate mgf files (mgf generators are publically available on the web). For interpeptide crosslinker analysis, high mass measurement accuracy of the precursor and product ions is required (typically between 10-15 ppm) as well as the known mass of the modifications remaining on the labeled residues due to the nature of the cleavable crosslinker (which are listed below for SuDP in Section 3.6.2). Due to the asymmetric nature of the our CID-CXL-MS/MS crosslinkers, each dataset needs to be

1 searched twice in order to account for the combination of crosslinker modifications on both
2 peptides of interpeptide crosslinked species since a given reactive residue side chain of a protein
3 can react with either end of our CID-CXL-MS/MS reagents. It should also be noted that these
4 asymmetric interpeptide crosslink pairs may be chromatographically resolved or essentially
5 coelute; although we have not conducted a formalized systematic study, the degree of separation
6 or coelution is dependent upon the type of stationary phase used in the reversed-phase separation
7 and the amino acid residue sequence and length of the peptides involved in the crosslink.

9 ***3.6.2 Analysis of crosslinked products using data obtained from LC/MS³ acquisitions***

10 All product ion spectra resulting from LC/MS³ acquisitions can be searched using any
11 database or de novo searching algorithm that is used to identify modified linear peptides as
12 previously described [4, 22, 23, 27, 29] with LC/MS² data being used for the identification of
13 deadend and intrapeptide crosslinks and LC/MS³ data for interpeptide crosslinks. Instead of
14 providing specific details for a particular search engine, parameters and searching criteria
15 important in any database or de novo search using peptides modified by our SuDP crosslinker
16 are presented. A fixed modification for carbamidomethylated Cys (57.0215 u) and variable
17 modifications for oxidized Met (15.9949 u), succinimidyl aspartate (SuD)-modified Lys
18 (197.0324 u), proline (P)-modified Lys (97.0528 u), SuDP deadend modified Lys (294.0852 u)
19 are applied for proteome-wide database searching. Additional variable modifications on His,
20 Ser, Thr, Tyr, and the N-terminus of the protein for the SuD (197.0324 u) and P (97.0528 u)
21 modifications along with semi-tryptic cleavage (i.e., a tryptic cleavage site at either end of the
22 peptide) can be applied to restricted database searching using the sequence of the protein of
23 interest in order to find very rare crosslinks that form due to the unique pK_a lowering

environments present on proteins [34]. The scoring cut off threshold needs to be the same as used for scoring modified peptides from proteomic analyses, such as an E-value or q-value of 0.01 or less. To identify interpeptide crosslinks, product ion spectra above the peptide identification threshold are arranged according to their retention time (i.e., scan numbers). These sorted results are searched for a product ion (MS^3) spectrum that identifies the presence of a peptide with a SuD mass modification that was within a retention time window (i.e., several consecutive MS^3 scans) with another product ion (MS^3) spectrum for a peptide containing a P mass modification. Although relatively abundant interpeptide crosslinks may take up to 60 s for the asymmetric pair to completely elute, most retention times are much smaller; in any case, the SuD- and P-modified peptides are often identified within consecutive MS^3 scans based on the ions produced in the previous MS^2 scan event of the interpeptide crosslink [23, 27, 29]. For these peptide pairs, the precursor ion present in the survey MS spectrum (the first MS scan event) is inspected for the presence of the calculated m/z of the intact interpeptide crosslink. The following product ion spectrum (the second MS scan event) is also examined for the presence of the two individual SuD- and P-modified peptide product ions generated from dissociation at the aspartyl-prolyl bond within the crosslinking reagent in order to verify that the modified peptides identified from the MS^3 product ion spectra were unequivocally originated from the CID-CXL-MS/MS interpeptide crosslink. This comparison can be achieved by manual verification or implementing in-house developed scripts as previously described [29]. It should be noted that the nature of the acquired data is readily amenable to custom-based scripts that can be written to facilitate the process of automating CID-CXL-MS/MS data analysis in any lab specific data management system and accompanying workflows.

4. Results and Discussion

For structural analysis of proteins and protein-protein interactions, the conditions required for promoting effective crosslinking needs to be established. The methods and protocols for our CID-CXL-MS/MS approach can be used as described in Section 3, but some aspects are worthy of discussion in order to address the analytical aspects of the crosslinking reaction and some additional alternatives for identification of interpeptide crosslinks.

4.1 Establishing and assessing effective protein crosslinking reaction conditions

The CID-CXL-MS/MS reagents are homobifunctional crosslinkers containing amine reactive N-hydroxysuccinimide (NHS) esters (Figure 2), thus effective crosslinking conditions for a given protein or protein complex and appropriate pH conditions to promote protein solubility and crosslinker reactivity need to be established. For reagents like SuDP, as well as any other crosslinkers that utilize NHS-esters, the optimal amine reactivity is going to be around pH 7.5 (i.e., between pH 7-8). When the pH is lower than 7, the kinetics of the crosslinking reaction is reduced, and above pH 8, hydrolysis of the NHS ester becomes an issue. For proteins having acidic or basic pIs, this is not an issue, but if the protein has a pI close to 7, then a pH at 6 or 8 is needed and will compromise labeling using crosslinking reagents with NHS or sulfo-NHS esters. Another factor to consider is buffer compatibility. Phosphate buffer is preferred to due to its ability to buffer at neutral pH and being unreactive to NHS-based crosslinkers. If a protein prefers amine-based buffers, those containing tertiary amines such as MOPS or HEPES are acceptable, unlike Tris which contains a primary amine and thus is often use to quench the crosslinking reaction.

1 In developing the use of our CID-CXL-MS/MS reagents, assessment of various reaction
2 conditions can be performed with a variety of biochemical methods. One that is easily
3 implemented uses SDS PAGE analysis to monitor the formation of crosslinked proteins and
4 protein complexes as a function of increasing crosslinker concentrations. This type of analysis
5 provides a visual confirmation whether suitable crosslinking or over-crosslinking is occurring
6 based on the increased detection of higher order oligomers and the decrease in the presence of
7 lower molecule weight monomers. If the stoichiometric oligomerization state of a protein is
8 unknown, it must be determined or verified by other biochemical measurements since
9 identification of crosslinked peptides originating from over-crosslinked samples may not reflect
10 native structural features or interactions. After crosslinking, confirmation of secondary structural
11 integrity can be assessed via circular dichroism as we have performed previously[27], whereas
12 potential changes in tertiary structure can be probed by conducting functional or enzymatic
13 assays. Typically a protein-to-crosslinker ratio of 1:100 is used with an incubation time of 30
14 min at ambient temperature and suitable for providing sufficient crosslinking without altering
15 protein structure of the average cytosolic protein, but some other method of verification to assess
16 protein conformation dynamics or regions of high flexibility may be required to explain
17 crosslinks that do not conform to a defined maximum distance constraint for a given structure or
18 structural model.

19 Since there may not be a suitable assay to measure secondary or tertiary structural
20 changes due to the protein complex being studied, we have started to implement the labeling of
21 the proteins with acetyl-NHS (Figure 3), which has the same reactive moiety as our CID-CXL-
22 MS/MS reagents, but its smaller size allows for all potential reactive Lys residues under the
23 given crosslinking conditions to be readily measured by LC/MS/MS analysis. These data can be

used to assess potential Lys residue amine reactivity with the NHS-based CID-CXL-MS/MS reagents, and the application of different proteases during digestion can be used to generate a peptide map to aid in the selection of a suitable proteases or proteases for the digestion of crosslinked samples. In addition the use of noncleavable crosslinkers like DSS, which is very similar in length to SuDP and SuVP and has the same reactive ester moieties, could be used to test sample labeling and workup conditions prior to using the synthesized reagents.

4.2 Assessing SCX enrichment of interpeptide protein crosslinks

Although the detection of crosslinked products can be performed from digested proteins labeled with our CID-CXL-MS/MS reagents without implementing SCX enrichment, fractionation using offline SCX facilitates the enrichment of higher charged state peptides, thus increasing the detection of interpeptide crosslinks. The SCX enrichment described in Section 3.2 utilizes a series of elutions containing increasing concentrations of NaCl (up to 1 M), but this required each fraction to be desalted using SPE prior to LC/MS/MS analysis. We have currently modified this method using ammonium formate since it is a volatile buffer and is easily removed during solvent evaporation, thus obviating the need for SPE for each collected fraction. To eliminate losses during sample handling, we have also used online SCX by modifying a nanoLC system by adding an in-house packed SCX column coupled to a reversed-phase trap column attached to a reversed-phase nanobore separation column; although this approach is relatively effective, more work is required to provide reproducible separations. Regardless of the SCX variation used, the efficiency of interpeptide crosslink detection is limited by missed cleavages in unmodified peptides, which lead to higher charge state containing peptides that co-elute with

crosslinked peptides. As a result, even with SCX enrichment, multiple LC/MS² and LC/MS³ analysis of collected fractions is often needed to characterize the interpeptide crosslinks.

4.3 Integrating LC/MS² and LC/MS³ data acquisition and analysis for confident CID-CXL-MS/MS interpeptide crosslink identification

On the basis of the design of our CID-CXL-MS/MS reagents, crosslinked products can be identified by utilizing LC/MS² and/or LC/MS³ acquisitions as illustrated in Figure 4. As demonstrated in our previous work using only a Thermo LCQ Deca ion trap [22, 23, 27], LC/MS³ analysis is well-suited for ion trap mass spectrometers where both deadend and intrapeptide crosslinked products are identified in the product ions generated by MS² CID, while the two released peptides of interpeptide crosslinks are individually fragmented and identified by product ions generated by MS³ CID. The LC/MS³ acquisition permits individual peptide identification and site of crosslink modification of the CID-CXL/MS/MS interpeptide crosslink via database searching or de novo sequencing, which is then correlated to the precursor ion fragmented at the MS² CID event, thus successfully identifying the interpeptide crosslink. Initially we demonstrated that changing voltages in the in-source region of an instrument can facilitate cleavage of the crosslinker and permit the released peptides to be analyzed during the MS² scan event [22], thus in theory, enabling any tandem mass spectrometer to identify CID-CXL-MS/MS interpeptide crosslinks.

More recent protein structural studies using the LC/MS³ acquisition approach with our CID-CXL-MS/MS reagents have involved hybrid ion trap-ICR [29] and ion trap-orbitrap systems [4] since they are capable of multistage MSⁿ acquisitions and performing measurements with high mass measurement accuracy and high resolving power which aid in confident

identification. Further characterization of product ions generated from CID-CXL-MS/MS interpeptide crosslinks at the MS² CID (or HCD) stage using these instruments revealed the presence of detectable and informative peptide backbone product ions. These ions were of sufficient number and detectable abundance (based on the peptides comprising the interpeptide crosslink) but did not preclude the formation of the released linear peptides that can be subjected to MS³ CID events as shown in the MS² product ion spectra reported in our previous work [4, 29] and illustrated in the data presented in Figure 5. The BSA interpeptide crosslink shown in Figure 5 was also identified in our previous study[29]in which a combination of an LC/MS² acquisition followed by a subsequent targeted LC/MS³ analysis was shown to identify 28 unique BSA interpeptide crosslinks using an unrestricted protein database. Based on the crystallographic data used for structural interpretation, the interpeptide crosslink shown was present on Lys residues within the maximum distance constraint of 23.9Å for the SuDP crosslinker. As a result, instruments such as Q-TOFs and orbitrap Q-Exactives can also be used to generate informative LC/MS² and LC/MS³ data when performing protein crosslinking using our CID-CXL-MS/MS reagents.

When using hybrid mass analyzers, the collected LC/MS² data can be analyzed by database searching or de novo sequencing algorithms for the identification of CID-CXL-MS/MS deadend and intrapeptide crosslinks (when acquiring product ion data for charge states $\geq +2$) or analyzed by searching algorithms designed for interpeptide crosslinks (when acquiring product ion data for charge states $\geq +4$). As indicated in Section 3.6.1, we have used MeroX [20], a publically available search engine with a user-friendly interface, to identify CID-CXL-MS/MS interpeptide crosslinks. Due to the asymmetric nature of our crosslinkers, each released linear peptide can exist in two forms, such as the SuD- and P-modified species if the SuDP crosslinker

is used for labeling (Figure 5), thus two searches are required to identify both interpeptide species. However, only one form of the interpeptide crosslink may be preferentially abundant due to the chemical and topological nature of the protein being labeled. To take advantage of the asymmetry of our crosslinkers, we are finalizing a searching algorithm (called CXLink) to better identify interpeptide CID-CXL-MS/MS crosslinks using unrestricted databases with statistical significance using data solely obtained from LC/MS² acquisitions. Nevertheless, we continue to assess the various search engines for interpeptide crosslinks that become available and encourage others to utilize this strategy when implementing our CID-CXL-MS/MS approach (or other crosslinking approaches). According to our limited surveys we have concluded that for confident structural analysis, manual inspection of identified interpeptide crosslinks is required for verification in order to establish a threshold cutoff score for confident identification for a given search algorithm.

On the basis of the CID-CXL-MS/MS interpeptide crosslinks subjected to either CID or HCD fragmentation, the length of each peptide present within the crosslink and the sequence position of the modified residue have an effect on the number and abundance of detectable product ions. For cases where the two released linear peptides are clearly detected but lack a sufficient number of product ion needed for confident identification, a targeted LC/MS³ approach can be used as we previously described [29]. In this type of LC/MS³ experiment, acquisitions can be performed using hybrid ion trap-ICRs and ion trap-orbitraps as well as other hybrid systems capable of MS³ scanning events like AB SCIEX QTRAPs and TripleTOFs by implementing an inclusion list containing MS m/z and MS² m/z values based on the data obtained from the LC/MS² data set. Although this targeted approach requires more sample and additional instrument time, it does provide an alternative method to identify and/or verify poor assignments

obtained for CID-CXL-MS/MS interpeptide crosslinks obtained from only LC/MS² data set acquisitions. If an instrument is not capable of LC/MS³ acquisitions, a higher collision energy setting during CID/HCD may be used in a targeted or non-targeted approach in order to provide more detectable ions of CID-CXL-MS/MS interpeptide crosslinks during the MS² CID/HCD event.

5. Conclusions

At this point in our research, we have laid the foundation for developing a comprehensive and potentially high-throughput mass spectrometry method for studying protein structural dynamics, interactions, and interaction networks using gas-phase cleavable crosslinkers. We have presented here the methods and protocols needed to successfully use our CID-CXL-MS/MS reagents in structural analysis of protein and protein interactions and have provided some alternative considerations for determining appropriate crosslinking reaction conditions, interpeptide crosslinking enrichment, and mass spectrometry data acquisitions and post-acquisition analyses to customize an approach for a given protein study and available LC-MS/MS analytical system. Although not discussed, our CID-CXL-MS/MS approach is also amenable to combinatorial generation of a library of crosslinkers by inserting various amino acids at the X position of the SuDX linker region, introduction of affinity tags for enrichment of crosslinked products, the utilization of stable isotope coding of reagents to enable relative quantitation and the analytical benefits that may be afforded by the use of ion mobility measurements to enhanced detection crosslinked products in order to capture the heterogeneity of protein structural interactions and dynamics.

Acknowledgements

The authors thank the research agencies of North Carolina State University and the North Carolina Agricultural Research Service for continued support of biological mass spectrometry research. This work was supported by grants from the National Science Foundation (DBI-1126244 and MCB-1021363).

References

- [1] E.V. Petrotchenko, C.H. Borchers, Crosslinking combined with mass spectrometry for structural proteomics, *Mass spectrometry reviews*, 29 (2010) 862-876.
- [2] A. Sinz, Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions, *Mass spectrometry reviews*, 25 (2006) 663-682.
- [3] J. Rappsilber, The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes, *Journal of structural biology*, 173 (2011) 530-540.
- [4] A.L. Olson, F. Liu, A.T. Tucker, M.B. Goshe, J. Cavanagh, Chemical crosslinking and LC/MS analysis to determine protein domain orientation: application to AbrB, *Biochemical and biophysical research communications*, 431 (2013) 253-257.
- [5] B.J. Greber, D. Boehringer, M. Leibundgut, P. Bieri, A. Leitner, N. Schmitz, R. Aebersold, N. Ban, The complete structure of the large subunit of the mammalian mitochondrial ribosome, *Nature*, 515 (2014) 283-286.
- [6] C. Plaschka, L. Lariviere, L. Wenzek, M. Seizl, M. Hemann, D. Tegunov, E.V. Petrotchenko, C.H. Borchers, W. Baumeister, F. Herzog, E. Villa, P. Cramer, Architecture of the RNA polymerase II-Mediator core initiation complex, *Nature*, 518 (2015) 376-380.
- [7] Q. Zhang, E. Crosland, D. Fabris, Nested Arg-specific bifunctional crosslinkers for MS-based structural analysis of proteins and protein assemblies, *Analytica chimica acta*, 627 (2008) 117-128.
- [8] S. Peri, H. Steen, A. Pandey, GPMAW--a software tool for analyzing proteins and peptides, *Trends in biochemical sciences*, 26 (2001) 687-689.
- [9] B. Schilling, R.H. Row, B.W. Gibson, X. Guo, M.M. Young, MS2Assign, automated assignment and nomenclature of tandem mass spectra of chemically crosslinked peptides, *Journal of the American Society for Mass Spectrometry*, 14 (2003) 834-850.

- 1 [10] S. Wefing, V. Schnaible, D. Hoffmann, SearchXLinks. A program for the identification of
2 disulfide bonds in proteins from mass spectra, *Analytical chemistry*, 78 (2006) 1235-1241.
- 3 [11] L.J. de Koning, P.T. Kasper, J.W. Back, M.A. Nessen, F. Vanrobaeys, J. Van Beeumen, E.
4 Gherardi, C.G. de Koster, L. de Jong, Computer-assisted mass spectrometric analysis of naturally
5 occurring and artificially introduced cross-links in proteins and protein complexes, *The FEBS*
6 *journal*, 273 (2006) 281-291.
- 7 [12] O. Rinner, J. Seebacher, T. Walzthoeni, L.N. Mueller, M. Beck, A. Schmidt, M. Mueller, R.
8 Aebersold, Identification of cross-linked peptides from large sequence databases, *Nature*
9 *methods*, 5 (2008) 315-318.
- 10 [13] S. Choi, J. Jeong, S. Na, H.S. Lee, H.Y. Kim, K.J. Lee, E. Paek, New algorithm for the
11 identification of intact disulfide linkages based on fragmentation characteristics in tandem mass
12 spectra, *Journal of proteome research*, 9 (2010) 626-635.
- 13 [14] S. McIlwain, P. Draghicescu, P. Singh, D.R. Goodlett, W.S. Noble, Detecting cross-linked
14 peptides by searching against a database of cross-linked peptide pairs, *Journal of proteome*
15 *research*, 9 (2010) 2488-2495.
- 16 [15] B. Yang, Y.J. Wu, M. Zhu, S.B. Fan, J. Lin, K. Zhang, S. Li, H. Chi, Y.X. Li, H.F. Chen,
17 S.K. Luo, Y.H. Ding, L.H. Wang, Z. Hao, L.Y. Xiu, S. Chen, K. Ye, S.M. He, M.Q. Dong,
18 Identification of cross-linked peptides from complex samples, *Nature methods*, 9 (2012) 904-
19 906.
- 20 [16] T. Walzthoeni, M. Claassen, A. Leitner, F. Herzog, S. Bohn, F. Forster, M. Beck, R.
21 Aebersold, False discovery rate estimation for cross-linked peptides identified by mass
22 spectrometry, *Nature methods*, 9 (2012) 901-903.
- 23 [17] M. Gotze, J. Pettelkau, S. Schaks, K. Bosse, C.H. Ihling, F. Krauth, R. Fritzsche, U. Kuhn,
24 A. Sinz, StavroX--a software for analyzing crosslinked products in protein interaction studies,
25 *Journal of the American Society for Mass Spectrometry*, 23 (2012) 76-87.
- 26 [18] E.V. Petrotchenko, K.A. Makepeace, C.H. Borchers, DXMSMS Match Program for
27 Automated Analysis of LC-MS/MS Data Obtained Using Isotopically Coded CID-Cleavable
28 Cross-Linking Reagents, *Current protocols in bioinformatics / editorial board*, Andreas D.
29 Baxeavanis ... [et al.], 48 (2014) 8 18 11-18 18 19.
- 30 [19] E.V. Petrotchenko, J.J. Serpa, K.A. Makepeace, N.I. Brodie, C.H. Borchers, (14)N(15)N
31 DXMSMS Match program for the automated analysis of LC/ESI-MS/MS crosslinking data from
32 experiments using (15)N metabolically labeled proteins, *Journal of proteomics*, 109 (2014) 104-
33 110.
- 34 [20] M. Gotze, J. Pettelkau, R. Fritzsche, C.H. Ihling, M. Schafer, A. Sinz, Automated
35 assignment of MS/MS cleavable cross-links in protein 3D-structure analysis, *Journal of the*
36 *American Society for Mass Spectrometry*, 26 (2015) 83-97.

- [21] D.B. Lima, T.B. de Lima, T.S. Balbuena, A.G. Neves-Ferreira, V.C. Barbosa, F.C. Gozzo, P.C. Carvalho, SIM-XL: A powerful and user-friendly tool for peptide cross-linking analysis, *Journal of proteomics*, (2015).
- [22] E.J. Soderblom, M.B. Goshe, Collision-induced dissociative chemical cross-linking reagents and methodology: Applications to protein structural characterization using tandem mass spectrometry analysis, *Analytical chemistry*, 78 (2006) 8059-8068.
- [23] E.J. Soderblom, B.G. Bobay, J. Cavanagh, M.B. Goshe, Tandem mass spectrometry acquisition approaches to enhance identification of protein-protein interactions using low-energy collision-induced dissociative chemical crosslinking reagents, *Rapid communications in mass spectrometry : RCM*, 21 (2007) 3395-3408.
- [24] H. Zhang, X. Tang, G.R. Munske, N. Zakharova, L. Yang, C. Zheng, M.A. Wolff, N. Tolic, G.A. Anderson, L. Shi, M.J. Marshall, J.K. Fredrickson, J.E. Bruce, In vivo identification of the outer membrane protein OmcA-MtrC interaction network in *Shewanella oneidensis* MR-1 cells using novel hydrophobic chemical cross-linkers, *Journal of proteome research*, 7 (2008) 1712-1720.
- [25] H. Zhang, X. Tang, G.R. Munske, N. Tolic, G.A. Anderson, J.E. Bruce, Identification of protein-protein interactions and topologies in living cells with chemical cross-linking and mass spectrometry, *Molecular & cellular proteomics : MCP*, 8 (2009) 409-420.
- [26] F. Dreier, M.Q. Muller, A. Sinz, M. Schafer, Collision-induced dissociative chemical cross-linking reagent for protein structure characterization: applied Edman chemistry in the gas phase, *Journal of mass spectrometry : JMS*, 45 (2010) 178-189.
- [27] F. Liu, M.B. Goshe, Combinatorial electrostatic collision-induced dissociative chemical cross-linking reagents for probing protein surface topology, *Analytical chemistry*, 82 (2010) 6215-6223.
- [28] E.V. Petrotchenko, J.J. Serpa, C.H. Borchers, An isotopically coded CID-cleavable biotinylated cross-linker for structural proteomics, *Molecular & cellular proteomics : MCP*, 10 (2011) M110 001420.
- [29] F. Liu, C. Wu, J.V. Sweedler, M.B. Goshe, An enhanced protein crosslink identification strategy using CID-cleavable chemical crosslinkers and LC/MS(n) analysis, *Proteomics*, 12 (2012) 401-405.
- [30] T. Miron, M. Wilchek, A spectrophotometric assay for soluble and immobilized N-hydroxysuccinimide esters, *Analytical biochemistry*, 126 (1982) 433-435.
- [31] E.V. Petrotchenko, J.J. Serpa, D.B. Hardie, M. Berjanskii, B.P. Suriyamongkol, D.S. Wishart, C.H. Borchers, Use of proteinase K nonspecific digestion for selective and comprehensive identification of interpeptide cross-links: application to prion proteins, *Molecular & cellular proteomics : MCP*, 11 (2012) M111 013524.

- [32] M. Wilm, A. Shevchenko, T. Houthaeve, S. Breit, L. Schweigerer, T. Fotsis, M. Mann, Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry, *Nature*, 379 (1996) 466-469.
- [33] J.R. Wisniewski, A. Zougman, N. Nagaraj, M. Mann, Universal sample preparation method for proteome analysis, *Nature methods*, 6 (2009) 359-362.
- [34] S. Madler, C. Bich, D. Touboul, R. Zenobi, Chemical cross-linking with NHS esters: a systematic study on amino acid reactivities, *Journal of mass spectrometry : JMS*, 44 (2009) 694-706.

Figure Legends

Figure 1. The major types of peptide crosslinks produced from chemically crosslinked proteins. After proteolytic digestion of chemically crosslinked proteins, four major products are produced. The most abundant products are the unmodified peptides (non-crosslinked), followed by deadend, intrapeptide, and interpeptide crosslinked products. The blue and red lines represent peptides that can either originate from the same protein or two different proteins. The dotted black line represents the linker region of the crosslinker, and the black filled-in circle represents a gas-phase cleavable moiety within the linker region of the crosslink.

Figure 2. The CID-CXL-MS/MS reagents. The three CID-CXL/MS/MS reagents that we have synthesized and used for protein structural studies are shown. The distances provided are for the extended conformation of the linker region of the crosslinker. This linker region distance is used for determining a maximum distance constraint as shown. For crosslinking two Lys residue side chains with SuDP/SuVP or SuDPG, the calculated maximum distance constraint is 23.9Å or 27.8 Å, respectively.

Figure 3. General work-flow for applying CID-CXL/MS/MS chemical crosslinkers to the characterization of protein structure and protein complexes.

Figure 4. Data-dependent and targeted acquisitions implementing multidimensional activation for CID-CXL/MS/MS interpeptide crosslink analysis. (A) The LC/MS² and/or LC/MS³ acquisition strategies for CID-CXL/MS/MS interpeptide crosslink identification. Using the SuDP crosslink as an example, the SuDP interpeptide crosslink selected for CID at the MS² stage of the acquisition generates product ions that include the SuD- and P-modified linear peptides which can be selected in a data-dependent manner (based on abundance) or a targeted approach (based on an inclusion list) for subsequent CID at the MS³ stage. Search algorithms to identify peptides are used to identify the SuD- and P-modified linear peptides and the modified residue using the MS³ product ion spectra. Search algorithms to identify interpeptide crosslinks containing gas-phase cleavable moieties in their linker region, such as MeroX [20], are used to identify the peptides and the modified residues of the interpeptide crosslink using the MS² product ion spectra. As is sometime the case, the identification of the CID-CXL-MS/MS interpeptide crosslink may not be of high confidence based solely on the product ions generated as the MS² stage, so a subsequent analysis using a targeted LC/MS³ approach is implemented to individually identify the SuD- and P-modified linear peptides. (B) The detailed mechanism of the gas-phase cleavage of the aspartyl-prolyl bond of the SuDP linker region of an interpeptide crosslink to produce the SuD- and P-modified linear peptides. Although the exact origin of the proton that protonates the prolyl residue has not been confirmed, it is shown as being abstracted from the carboxylic acid of the aspartyl side chain which is consistent with the resulting mass

modifications on each generated linear peptide. The blue and red lines represent peptides that can either originate from the same protein or two different proteins.

Figure 5. LC/MS² and LC/MS³ fragmentation of an SuDP interpeptide crosslink for crosslinked BSA. The (A) MS spectrum shows the intact +5 interpeptide crosslink precursor which exists in two forms, $[M_1\text{-SuDP-}M_2+5H]^{5+}$ and $[M_2\text{-SuDP-}M_1+5H]^{5+}$ due to the asymmetry of the SuDP crosslink, that essentially coelute during the reversed-phase LC separation. The SuDP interpeptide crosslink precursor ion was selected as the most abundant eligible ion and fragmented during the MS² CID event to promote aspartyl-prolyl bond dissociation and produce four distinct linear peptides, $[M_1\text{-SuD}+2H]^{2+}$, $[M_2\text{-SuD}+2H]^{2+}$, $[P\text{-}M_1+2H]^{2+}$ and $[P\text{-}M_2+3H]^{3+}$, as the predominant ions in the (B) MS² product ion spectrum, but other informative product ions are also clearly present. The (C) and (D) annotated MS³ product ion spectra for linear peptides $[P\text{-}M_1+2H]^{2+}$ and $[M_2\text{-SuD}+2H]^{2+}$, respectively, are shown. The amino acid residue sequences for the M₁ and M₂ peptides are indicated with C* denoting carbamidomethylated cysteinyl residues.

Figure 1

Disuccinimidyl-succinamyl-proline (SuDP)

Disuccinimidyl-succinamyl-valyl-proline (SuVP)

Disuccinimidyl-succinamyl-aspartyl-prolyl-glycine (SuDPG)

Figure 2

Figure 3

Figure 4

Figure 5