

Optimized cross-linking mass spectrometry for in situ interaction proteomics

Zheng Ser^{1,2}, Paolo Cifani¹, Alex Kentsis^{1,2,3,4*}

¹ Molecular Pharmacology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY

² Tri-Institutional PhD Program in Chemical Biology, Memorial Sloan Kettering Cancer Center, New York, NY

³ Department of Pediatrics, Memorial Sloan Kettering Cancer Center, New York, NY

⁴ Department of Pediatrics, Pharmacology, and Physiology & Biophysics, Weill Cornell Medical College, Cornell University, New York, NY

*To whom correspondence should be addressed. Email: kentsisresearchgroup@gmail.com

Keywords

cross-linking, mass spectrometry, chromatin, BSA, protein-protein interactions, target-decoy strategy, database search, false positive discovery, proteomics

Abstract

Recent development of mass spectrometer cleavable protein cross-linkers and algorithms for their spectral identification now permits large-scale cross-linking mass spectrometry (XL-MS). Here, we optimized the use of cleavable disuccinimidyl sulfoxide (DSSO) cross-linker for labeling native protein complexes in live human cells. We applied a generalized linear mixture model to calibrate cross-link peptide-spectra matching (CSM) scores to control the sensitivity and specificity of large-scale XL-MS. Using specific CSM score thresholds to control the false discovery rate, we found that higher-energy collisional dissociation (HCD) and electron transfer dissociation (ETD) can both be effective for large-scale XL-MS protein interaction mapping. We found that the density and coverage of protein-protein interaction maps can be significantly improved through the use of multiple proteases. In addition, the use of sample-specific search databases can be used to improve the specificity of cross-linked peptide spectral matching. Application of this approach to human chromatin labeled in live cells recapitulated known and revealed new protein interactions of nucleosomes and other chromatin-associated complexes in situ. This optimized approach for mapping native protein interactions should be useful for a wide range of biological problems.

Introduction

Protein-protein interactions mediate cellular functions including signaling, metabolism and cell differentiation. Disruption of protein-protein interactions can contribute to disease and defining how these interactions change across the proteome can provide important insights into pathogenic mechanisms and ultimately improved therapies.¹ Conventional methods to study protein-protein interactions involve large-scale affinity purification or separation of protein complexes combined with protein identification using mass spectrometry.²⁻⁷ These methods typically require ectopic expression of proteins, affinity tags for purification or separation of protein complexes from extracts and other non-physiological conditions, which can produce artifacts and disruption of native protein interactions. In contrast, cross-linking mass spectrometry (XL-MS) relies on covalent labeling and cross-linking of proteins in physical proximity, thereby identifying physiologic protein-protein interactions. In addition, chemical cross-linking preserves direct or proximal protein-protein interactions, and can also identify transient interactions that occur dynamically during specific cell or developmental states. In addition, high-resolution mass spectrometry can identify direct sites of interaction between cross-linked amino acids, providing detailed spatial information for structural studies.⁸⁻⁹ Recently, development of collision-induced dissociation (CID) cleavable cross-linkers¹⁰⁻¹⁴ and improvements in algorithms for cross-link peptide spectral analysis¹⁵⁻¹⁹ has enabled protein-protein interaction mapping of complex proteomes, including worm,²⁰⁻²¹ bacteria,²¹⁻²³ and human^{15-16, 24-25} and various other organisms,²⁶ with hundreds-to-thousands of cross-links and protein-protein interactions.^{16, 21} While cross-linking of cell extracts provides direct control of the reaction to label specific protein complexes,^{15-16, 20-21} membrane-permeability of some cross-

linking reagents enables cross-linking of protein complexes in live cells to define specific protein-protein interactions *in situ*.

To facilitate large-scale interaction proteomics, we sought to define the optimal methods for the isolation of cross-linked peptides from protein complexes labeled *in situ*, and to ensure their accurate identification using a target-decoy strategy adapted for cross-linking mass spectrometry. Thus, we present optimized parameters for the use of the CID-cleavable cross-linker disuccinimidyl sulfoxide (DSSO) *in situ*. We found that higher-energy collisional dissociation (HCD) fragmentation produced as many cross-linked protein complex identifications as electron transfer dissociation (ETD) fragmentation using either whole proteome or focused target databases for cross-link peptide spectral matching. The use of multiple proteases significantly increased the coverage of protein-protein interactions maps. Importantly, the developed target-decoy strategy allowed for explicit control of false discovery and optimization of sensitivity and specificity of protein interactions maps in cells *in situ*. Application of this approach to human chromatin labeled in live cells recapitulated known and revealed new protein interactions of nucleosomes and other chromatin-associated complexes.

Experimental Section

Reagents. Disuccinimidyl sulfoxide (DSSO), formic acid, dimethyl sulfoxide (DMSO), and LC-MS grade solvents were obtained from Thermo Scientific. Bovine serum albumin (BSA), (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), sodium chloride, magnesium chloride, potassium chloride, sucrose, ethylenediaminetetraacetic acid (EDTA), dithiothreitol, guanidinium hydrochloride, *Staphylococcus aureus* micrococcal nuclease, iodoacetamide, ammonium bicarbonate and formic acid were obtained from Millipore Sigma. Protease inhibitors 4-(2-aminoethyl)-benzenesulfonyl fluoride hydrochloride (AEBSF) and pepstatin were obtained from Santa Cruz, bestatin from Alfa Aesar and leupeptin from EMD Millipore. Trypsin and GluC proteases were obtained from Promega. Chymotrypsin protease was obtained from Pierce Thermo. LysC was obtained from Wako Pure Chemical Industries.

Cell culture and protein isolation. HEK293T cells were cultured in DMEM supplemented with 10% (v/v) fetal bovine serum, penicillin and streptomycin. HEK293T cells (50 million) were sedimented by centrifugation at 500 g and lysed in 500 μ l 6M guanidine hydrochloride in 100 mM ammonium bicarbonate buffer, pH 8. The lysate was sonicated (E210 adaptive focused acoustic sonicator, Covaris) for 5 minutes at 4°C and homogenized by passing through a 27G needle. The lysate was then clarified by centrifugation at 18,000 g for 10 min at 4°C. Protein concentration of clarified lysates was determined using the bicinchoninic acid (BCA) assay, according to manufacturer's instructions (Thermo Scientific).

Preparation of cross-linked bovine serum albumin and peptide purification. 50 μ g of BSA was solubilized in 50 μ l HEPES buffer (20 mM HEPES, pH 8, 150 mM sodium chloride, 1.5 mM magnesium chloride, 0.5 mM dithiothreitol) and cross-linked by adding 1 μ l of 50 mM

DSSO dissolved in DMSO (1:1 protein amine:DSSO molar ratio). The solution was incubated at room temperature for 1 hour, before quenching by the addition of 1M ammonium bicarbonate buffer to a final concentration of 20 mM. The cross-linked protein was denatured by adding 50 μ l of 6 M guanidine hydrochloride in 100 mM ammonium bicarbonate, pH 8, then reduced by adding 11 μ l of 100 mM dithiothreitol at 56°C for 1hr and alkylated with 12 μ l of 550 mM iodoacetamide at room temperature for 30 min in the dark. Alkylation was quenched by the addition of 13 μ l of 100 mM dithiothreitol and incubated at 56°C for 20 min. The solution was diluted with 50 mM ammonium bicarbonate solution, pH 8, for a final concentration of 0.2 M guanidine hydrochloride. For digestion with a single protease, either trypsin, chymotrypsin or GluC was added in 1:50 (enzyme:protein, w/w) concentration and incubated at 37°C for 18 hours. For sequential digestion with LysC followed by trypsin, LysC was first added in 1:100 (enzyme:protein, w/w) concentration and incubated at 37°C for 8 hours. Trypsin was then added in 1:50 (enzyme:protein, w/w) concentration and incubated at 37°C for 18 hours. Digestion was stopped by the addition of formic acid to 1% (v/v) concentration. Peptides were then purified using solid phase extraction with C18 Macrospin columns according to manufacturer's instructions (Nest Group), and concentrated by vacuum centrifugation.

Protein cross-linking of live cells in situ and isolation of cross-linked peptides from chromatin. HEK293T cells (50 million) were sedimented by centrifugation at 500 g and then suspended in 1 ml of hypotonic buffer (10 mM HEPES, pH 8, 10 mM sodium chloride, 1 mM magnesium chloride, 0.5 mM dithiothreitol, protease inhibitors). For cross-linking, 50 μ l of 50 mM DSSO dissolved in DMSO was added and cells were incubated for 1 hour at 4°C. The reaction was then quenched by the addition of 50 μ l of 500 mM Tris-HCl, pH 8. Cells were subsequently subjected to 15 strokes of Dounce homogenization and the solution was then

centrifuged for 15 min at 3300 g to isolate nuclei. Sedimented nuclei were then resuspended in 500 µl of 0.25 M sucrose, 10 mM magnesium chloride buffer and layered on top of a solution of 500 µl of 0.88 M sucrose, 0.05 mM magnesium chloride. The resultant mixture was centrifuged for 10 min at 1200 g. Purified nuclei were then resuspended in 1 ml of 10 mM HEPES, pH 7.4, 0.2 mM magnesium chloride solution and incubated on ice for 10 min to extract nucleoplasm. The solution was then centrifuged for 20 min at 770 g. The pellet was next resuspended in 1 ml of 10 mM HEPES, pH 7.4, 500 mM sodium chloride, 0.2 mM magnesium chloride and incubated on ice for 10 min to extract bound proteins. The solution was then centrifuged for 20 min at 770 g. Next, the pellet was solubilized in 500 µl of detergent buffer (15 mM HEPES, pH 7.5, 500 mM sodium chloride, 1 mM dithiothreitol, 0.34 M sucrose, 10% glycerol, 0.5% v/v Triton X-100) and incubated on ice for 10 min. The solution was then layered on top of a 500 µl solution of 0.88 M sucrose, 0.05 mM magnesium chloride and centrifuged for 20 min at 3300 g to purify chromatin. The chromatin pellet was resuspended in 500 µl of 15 mM HEPES, pH 7.5, 15 mM sodium chloride, 60 mM potassium chloride, 5 mM magnesium chloride, 1 mM calcium chloride, 0.25 M sucrose and protease inhibitors (0.5 M AEBSF, 0.001 mM pepstatin, 0.01 mM bestatin and 0.1 mM leupeptin). The chromatin suspension was incubated at 37°C for 10 min. Five units of micrococcal nuclease were subsequently added and incubated at 37°C for 1 hour. The nuclease digestion was stopped by the addition of EDTA to a final concentration of 50 mM and the solution was centrifuged for 10 min at 18,000 g to remove the nuclear matrix. The supernatant containing cross-linked chromatin was then quantified using the BCA assay.

Cross-linked chromatin proteins were then purified by filter-aided sample preparation (FASP) using Sartorius Vivacon 500 with a 30,000 MW cut-off filter, as previously described.²⁷⁻²⁸ Briefly, 200 µg of protein was diluted 5-fold in urea buffer (8 M urea, 50 mM ammonium

bicarbonate, pH 8.0) and applied to the filter unit. Filter units were centrifuged at 14,000 *g* for 20 min. The membrane was then washed twice with 200 μ l urea buffer by centrifugation at 14,000 *g* for 20 min. The membrane was then incubated with 200 μ l of 100 mM dithiothreitol in urea buffer for 20 min at room temperature before centrifugation at 14,000 *g* for 20min. The proteins were then incubated with 200 μ l of 100 mM iodoacetamide in urea buffer for 20 min at room temperature in the dark before centrifugation at 14,000 *g* for 20 min. The membrane was washed twice more with urea buffer then once with 200 μ l of 50 mM ammonium bicarbonate, pH 8. Proteins retained on the membrane were then digested with trypsin at 1:50 (enzyme:protein, w/w) concentration at 37°C for 18 hours. Digested peptides were collected by centrifugation at 14,000 *g* for 20 min. Peptides were then purified using solid phase extraction with C18 Macrospin columns according to manufacturer's instructions (Nest Group). The purified peptides were concentrated by vacuum centrifugation and resuspended in 100 μ l of 5% acetonitrile, 0.1% formic acid. The peptides were then injected onto the strong cation exchange (SCX) column (Protein Pak Hi-Res SP 7 μ m 4.6 x 100mm, Water) at a flow rate of 0.5 ml/min with a column temperature of 30°C using the Alliance HPLC system (Waters). Gradient was run with 5% acetonitrile and 0.1% formic acid in water for solvent A and 1 M potassium chloride in 5% acetonitrile and 0.1% formic acid in water for solvent B. Column gradient was set as follows: 0-3 min (0% B); 3-33 min (0-10% B); 33-43 min (10-100% B); 43-60 min (100% B); 60-70 min (100-0% B); 70-90 min (0% B). Six fractions were collected from 38 to 60 min and purified using solid phase extraction with C18 Macrospin columns according to manufacturer's instructions (Nest Group). Peptides were concentrated by vacuum centrifugation and then stored at -20°C before analysis.

Liquid chromatography and nanoelectrospray mass spectrometry. Cross-linked peptides were separated by reverse phase nanoflow liquid chromatography (EKspert nanoLC 425, Eksigent) coupled to the Orbitrap Fusion mass spectrometer (Thermo). Cross-linked BSA peptides and HEK293T peptides were resuspended in 0.1% formic acid in water (v/v). For cross-linked BSA peptides, 1 μg of peptide was injected for analysis. For cross-linked BSA peptides mixed with non-cross-linked HEK293T peptides in 1:1 (w/w) ratio, 1 μg of total peptide was analyzed. For SCX fractions of cross-linked chromatin, peptides were resuspended in 20 μl of 0.1% formic acid in water (v/v) and 0.5 μg of peptides were analyzed.

For liquid chromatography, a trap-elute system was used, as described previously.²⁹ Reverse phase columns were fabricated as previously described.³⁰ Trap columns were fabricated by packing Poros R2 10 μm C18 particles (Life Technologies) into 4 cm fritted capillaries with internal diameters of 150 μm . Reverse phase columns were fabricated by packing Reprisil 1.9 μm silica C18 particles into 40 cm fritted capillaries with internal diameters of 75 μm . Peptides were resolved over a 90 min gradient from 5% to 40% acetonitrile and ionized using the DPV-565 Picoview ion source (New Objective) operated with the ionization voltage of 1700 V. Custom electrospray emitters were fabricated as previously described.³¹

For mass spectrometry, four different acquisition methods were used: 1) CID-MS2/HCD-MS2, 2) CID-MS2/ETD-MS2, 3) CID-MS2/EThcD-MS2, 4) CID-MS2/HCD-MS3, where HCD is the higher-energy collisional dissociation, ETD is electron transfer dissociation, and EThcD is the hybrid electron transfer with supplemental activation by higher energy collision dissociation. For MS2-MS2 methods, precursor ion spectra were recorded at 400-1800 m/z with 60,000 m/z Orbitrap resolution, automatic gain control target of 1×10^5 ions and maximum injection time of 50 ms. Precursor ions with 3-10 positive charge were selected for MS2 fragmentation with

dynamic exclusion of 60 secs after 1 scan and isolation window of 2 Th. For CID-MS2, precursor ion spectra were recorded in the Orbitrap with resolution 30,000 m/z, automatic gain control target of 5.0×10^4 and maximum injection time of 100 ms at normalized collision energy of 30%. For HCD-MS2, precursor ion spectra were recorded in the Orbitrap with resolution 30,000 m/z, automatic gain control target of 5.0×10^4 and maximum injection time of 120 ms at normalized collision energy of 30%. For ETD-MS2, precursor ion spectra were recorded in the Orbitrap with resolution 30,000 m/z, automatic gain control target of 5.0×10^4 ions and maximum injection time of 120 ms with fluoroanthene reaction time of 50 ms, ETD target of 2.0×10^5 ions and maximum fluoroanthene injection time of 200 ms. For EThcD-MS2, precursor ion spectra were recorded in the Orbitrap with resolution 30,000 m/z, automatic gain control target of 5.0×10^4 ions and maximum injection time of 120 ms with reaction time of 50 ms, reagent target of 2.0×10^4 and maximum fluoroanthene injection time of 200 ms and supplemental collision energy of 20%. For HCD-MS3, MS2 fragment ions with 2-6 charge states and mass difference of 31.9721 Da with 1-100% intensity range were selected for MS3 fragmentation, with a mass tolerance of 30 ppm and 2 Th isolation window. MS3 fragment ion spectra were recorded in the linear ion trap acquisition with automatic gain control target of 2.0×10^4 ions and maximum injection time of 150 ms.

Data analysis. Acquired mass spectra raw files were analyzed using XlinkX version 2.2,¹⁵⁻¹⁶ as implemented in Proteome Discoverer version 2.2. The target search database was the *Homo sapiens* proteome from SwissProt, containing isoforms³² (version January 2016), supplemented with the sequences of common contaminant proteins from cRAP.³³ Mass spectral analysis parameters included 2 maximum allowed missed cleavages, minimum peptide length of 5 and 4 maximum variable modifications: methionine oxidation, hydrolysis of lysyl-DSSO by water, and

lysyl-DSSO-Tris adduct. Precursor mass tolerance was set at 10 ppm with fragment mass tolerance of 20 ppm and 0.5 Da for Orbitrap and ion trap spectra, respectively. Cysteine carbamidomethylation was set as fixed modification. A 1% false discovery rate (FDR) was set for cross-link peptide spectra matching with minimum score threshold set at 0 using the Percolator algorithm.³⁴ Peptide spectral matches and their scores were analyzed using R scripts to concatenate cross-links identified from biological replicates (<https://github.com/kentsisresearchgroup/R-script-for-XLMS-data-analysis>). For histone proteins, peptides that could be identified as exclusively belonging to specific isoforms were marked with their specific isoform protein name or gene name. Proteins that could be mapped to multiple isoforms were labeled with the family protein or gene name, e.g. peptides that could belong to histone isoforms H3.1, H3.2 or H3.3 were labeled as histone H3 peptides. Bar graphs and scatter plots were plotted using Origin 2018 (Microcal). Venn diagrams were made with Venny version 2.1.³⁵ Peptide cross-link sequences were visualized using xiNET.³⁶ Atomic resolution structures were visualized with UCSF Chimera³⁷ version 1.12 and cross-links were mapped using Xlink analyzer.³⁸ Protein-protein interaction maps were made with Cytoscape version 3.6.0.³⁹ Mass spectrometry raw files and search results are publicly available through the ProteomeXchange data repository via the PRIDE database.

Results and Discussion

Calibration of cross-link peptide-spectra matching (CSM) scores to control sensitivity and specificity of large-scale XL-MS. MS-cleavable cross-linkers are designed to improve the accuracy and sensitivity of identification of protein-protein interactions by mass spectrometric analysis and subsequent statistical spectral matching. Specifically, the DSSO cross-linker is labile upon collision induced dissociation, leading to decoupling of cross-linked peptides into linear peptides in the gas-phase, and consequently enabling independent identification of each peptide. In addition, fragmentation of the sulfoxide cross-linker leaves characteristic adducts on the de-coupled peptides, thus producing diagnostic mass differences that can be leveraged to improve the specificity of identification of cross-linked peptide ions¹². Depending on the experimental parameters used, fragmentation of the peptide backbone can then be achieved by selecting either the cross-linked precursor ion or the decoupled peptide ions. Algorithms based on scoring functions for matching of cross-linked peptide spectra have recently been described.^{15-16, 19}

To further advance our ability to control the sensitivity and specificity of cross-linked peptide discovery, we reasoned that discrimination of true and false positive peptide-spectral matches can be achieved using target-decoy strategies tailored specifically to cross-linked peptides. To model a typical XL-MS sample, containing both cross-linked and non-cross-linked peptides, we cross-linked purified bovine serum albumin (BSA), which was subsequently diluted in non-cross-linked proteins extracted from human HEK293T cells (Figure 1A). BSA is a widely used model protein for mass spectrometry that spontaneously oligomerizes in solution.⁴⁰⁻⁴¹ Insofar as HEK293T cells do not express human albumin or proteins with sequence homology to BSA, this mixture model of true positive (cross-linked BSA) and true negative (human proteins) should

provide a faithful model to estimate and control the FDR of large-scale cross-linked mass spectrometry. Thus, we use a generalized linear mixture model to estimate the false discovery rate among four different methods of cross-linked peptide analysis: CID-MS2/HCD-MS2, CID-MS2/ETD-MS2, CID-MS2/EThcD-MS2 and CID-MS2/HCD-MS3 (Figure 1B). Despite applying a 1% FDR threshold for peptide-spectral matching using Percolator,³⁴ we observed a substantial number of apparently incorrectly matched cross-linked peptide spectra. For example, for CID-MS2/HCD-MS2 fragmentation, 173 of 291 (59%) of cross-links matched were mapped to human proteins that were not DSSO cross-linked, indicating incorrect CSMs. Manual inspection of representative spectra confirmed that these apparent incorrect matches were due to either incomplete fragmentation spectrum of one of the peptides in the pair or relatively low signal intensity of the diagnostic DSSO fragment ions. We confirmed that none of the examined spectra from apparently incorrect matches exhibited fragment ions consistent with albumin peptide sequences. Finally, we also verified that none of the mis-matched spectra were due to the incomplete DSSO cross-linking reaction due to spectra with single site DSSO adducts.

We reasoned that such false cross-linked peptide-spectral matches can be distinguished by their relative cross-link peptide-spectra matching (CSM) scores, as defined by the current XlinkX algorithm (Figure 1B).¹⁵⁻¹⁶ Thus, we fit the observed CSM score distributions to a linear mixture of two Gaussian functions, producing two modes corresponding to apparently true and false matches (Figure 1C). Consequently, for the observed spectra using the four acquisition methods tested, this led to a CSM score threshold of 100 to achieve 1% FDR at the peptide level (Figure 1D). In addition, we defined a CSM score threshold of 74 for improved sensitivity at a more relaxed (10%) peptide-level FDR. These empiric CSM score thresholds should be generalizable

to other experiments, insofar as the physicochemical properties of the peptide analyzed are adequately approximated by BSA and human proteins (Figures S-1).

Importantly, the use of these CSM score thresholds reduced the number of false non-BSA cross-links matched to non-cross-linked human proteins from 58 to 3 (Figure 1E). In addition, we analyzed the identified BSA cross-links with respect to the high-resolution atomic-level structure of BSA (PDB ID 4f5s).⁴² Consistently, the use of CSM score thresholds eliminated cross-links between lysine residues that are farther than 23 Å (the length of the DSSO cross-linker), while retaining cross-link matches with the expected physical distances (Figure 1F). These results demonstrate the need for controlling the false identifications inherent in the statistical spectral matching for large-scale cross-linking mass spectrometry proteomics, and provide a facile means to do so using a target-decoy spectral matching strategy adapted for cross-linking mass spectrometry.

Improved cross-linked peptide identification using CID-MS2/HCD-MS2 fragmentation.

Comparison of CID, HCD and ETD fragmentation of linear non-cross-linked peptides indicates that the efficiency of each fragmentation method, and therefore their capacity to produce optimal fragment ion spectra, depends on the specific physicochemical properties of fragmented peptides⁴³. For example, ETD exhibits reduced fragmentation efficiency for larger m/z peptides.⁴⁴ Previous studies of non-cleavable cross-linkers found that HCD fragmentation increased the identification of cross-linked peptides with charge states of 3-4, while the hybrid EThcD method facilitated the identification of cross-linked peptides with charge states of ≥ 5 .⁴⁵ Thus, to assess the various methods of identification of DSSO cross-linked peptides, we used the mixture of cross-linked BSA and non-cross-linked human proteome. We compared 4 different

fragmentation methods, with CID-MS2 used for cross-link fragmentation: CID-MS2/HCD-MS2, CID-MS2/ETD-MS2, CID-MS2/EThcD-MS2 and CID-MS2/HCD-MS3.

We observed an average of 43 cross-links per experiment across 3 biological replicates identified by CID-MS2/HCD-MS2, as compared to 25, 19 and 7 cross-links identified with CID-MS2/ETD-MS2, CID-MS2/HEThcD-MS2 and CID-MS2/HCDMS3 methods, respectively (Figure 2A). A total of 56 unique cross-links across 3 biological replicates were identified using CID-MS2/HCD-MS2 fragmentation, while 35 unique cross-links were identified from CID-MS2/ETD-MS2, of which only 3 unique cross-links identified by CID-MS2/ETD-MS2 were not identified by CID-MS2/HCD-MS2 (Figure 2B). For cross-linked BSA peptides mixed with the non-cross-linked proteome, an average of 21 cross-links per experiment across 3 biological replicates were identified by CID-MS2/HCD-MS2, as compared to 19, 13 and 3 cross-links identified by CID-MS2/ETD-MS2, CID-MS2/HEThcD-MS2 and CID-MS2/HCDMS3 methods, respectively (Figure 2C). 28 unique BSA cross-links across 3 biological replicates were identified from CID-MS2/HCD-MS2, while a comparable 27 unique cross-links were identified from CID-MS2/ETD-MS2. Both CID-MS2/HCD-MS2 and CID-MS2/ETD-MS2 identified 8 and 5 cross-linked peptides, not identified by other fragmentation methods, respectively, suggesting that there is some orthogonality in coverage between these two fragmentation methods (Figure 2D). Comparison between CID-MS2/HCD-MS2 and CID-MS2/ETD-MS2 methods showed an overlap between the types of cross-links identified, with 32 out of 59 (54%) and 19 out of 35 (54%) shared cross-links identified for cross-linked BSA peptides and cross-linked BSA peptides mixed with the non-cross-linked proteome, respectively. Comparison of identified cross-linked peptides showed relatively shorter peptide length, lower precursor m/z and lower cross-linked peptide identifications across all charge states obtained by ETD fragmentation

compared to that obtained by HCD fragmentation (Figure S-1), in line with prior observations.⁴⁴

A substantial number (10-30%) of cross-linked peptides were identified exclusively by either ETD or HCD fragmentation methods (Figure 2B, 2D), suggesting some orthogonality between the two fragmentation methods. Naturally, these results are generalizable to other complex proteomes largely for cross-linked peptides that are adequately modeled by cross-linked BSA.

In all, these results indicate that CID-MS2/HCD-MS2 fragmentation can yield the highest number of cross-links for simple samples, with a 60% increase in unique cross-link sites as compared to CID-MS2/ETD-MS2. For complex samples, CID-MS2/HCD-MS2 identified as many unique cross-link sites as CID-MS2/ETD-MS2. On the other hand, CID-MS2/HCD-MS3 and CID-MS2/EThcD-MS2 methods consistently recorded fewer cross-link spectra and fewer identifications when compared to CID-MS2/ETD-MS2 and CID-MS2/HCD-MS2 methods. This can be attributed to the relatively longer acquisition times required for MS3 and EThcD methods, as also noted in a recent study.¹⁶

Sample-specific search target databases for identification of protein-protein interactions by XL-MS. While the use of a mass spectrometer cleavable cross-linkers reduces the database search space and time as compared to non-cleavable cross-linkers,¹⁵ the use of sample-specific search databases can improve the sensitivity of mass spectral matching.⁴⁶⁻⁴⁹ Since cross-linking mass spectrometry is often used for the analysis of specific purified protein complexes, we examined the effects of using sample-specific as compared to proteome-wide search databases.

Thus, we generated search databases based on the BSA sequence alone, or a concatenated database containing BSA plus the human reference proteome, supplemented with common proteomic contaminants from cRAP.³³ We observed an average of 43 BSA cross-links across 3

biological replicates when searching against either the BSA-specific or reference human proteome databases (Figure 2E). For cross-linked BSA peptides mixed with the non-cross-linked proteome, we found an average of 21 and 22 BSA cross-links from 3 biological replicates when searching against the BSA-specific database or BSA plus reference human proteome databases, respectively (Figure 2E). This suggests that the database size and composition do not affect the sensitivity of cross-link identification *per se* (Figure 2F). With the BSA-specific database excluding any possible false matches, we observed a significant reduction in false matches of cross-linked BSA to non-cross-linked human proteins, when using the BSA-specific database, as would be expected from first principle considerations (0 as compared to 3, respectively, $p = 0.048$, Figure 2E).

Diverse sequence-specific proteases to expand the coverage and density of DSSO XL-MS protein interaction maps. We reasoned that the DSSO reaction with lysine residues reduces the accessibility of cross-linked peptides for protease digestion, thus reducing the identification of tryptic cross-linked peptides. Consistent with this, we found substantially more unique cross-links identified from the sequential protease digest using LysC endoprotease followed by trypsin, as compared to that of trypsin alone (90 versus 18, respectively, Figure 3A). This combined LysC and trypsin digestion produced cross-link sites across the whole BSA protein sequence (Figure 3C), suggesting that accessibility of amino acid sites for protease digestion affects cross-link peptide identification. For linear non cross-linked peptides, additional sequence-specific proteases are known to identify additional peptide sequences, thus improving protein sequence coverage.⁵⁰⁻⁵² Hence, we compared three different digestion methods using additional proteases specific to non-lysine amino acids: 1) sequential digestion with LysC followed by trypsin, 2) GluC endoprotease, and 3) chymotrypsin. Combined LysC and trypsin digestion identified the

highest number of unique cross-linked sites, as compared to GluC and chymotrypsin digestions (90 versus 69 and 56, respectively, $p < 0.01$ and $p < 0.01$, Figure 3B). We observed a minor overlap among these cross-linked sites, with the majority of identified sites involving different BSA residues (Figure 3C), presumably due to the differences in susceptibility to enzymatic endoproteolysis.⁵⁰ On average, cross-linked peptides identified by GluC digestion were more highly charged, as compared to those identified by chymotrypsin or trypsin (Figure S-2). Thus, the use of multiple, non-lysine-dependent proteases can expand the coverage and density of DSSO XL-MS protein interaction maps.

Native chromatin protein-protein interactions identified using cross-linking mass spectrometry *in situ*. Having identified a robust means to control the sensitivity and specificity of large-scale cross-linking mass spectrometry, we sought to examine the ability of DSSO to label native protein complexes in live human cells. We chose to study protein-protein interactions of human chromatin *in situ*, given its abundance and prominent functions in the regulation of cell growth and development. We optimized a protocol to label live human HEK293T cells with DSSO (see Methods). Chromatin proteins were extracted from purified nuclei by sucrose sedimentation, followed by sequential salt, detergent, and micrococcal nuclease digestion to release chromatin bound proteins.⁵³⁻⁵⁵ We used SDS-PAGE followed by silver staining to confirm the presence of reduced mobility protein complexes upon DSSO cross-linking, as compared to mock-treated cells (Figure 4A). Likewise, we used Western immunoblotting to verify the isolation of histone-containing chromatin complexes, as opposed to nucleoplasmic or cytoplasmic proteins, as assessed by histone H3, BRG1, and GAPDH, respectively (Figure 4B).

Analysis of the cross-linked chromatin fraction identified 1,277 unique cross-linked peptides from three biological replicates, with 384 unique cross-linked peptides at least two out of three biological replicates (Table S-1). Of these, 189 (49%) of cross-linked peptides were between two different proteins, while 195 (51%) were interactions within the same proteins, with 95 of these interactions also represented in the current BioGRID and CORUM protein-protein interaction databases.⁵⁶⁻⁵⁹ In agreement with the empirically determined CSM score thresholds for identifying high-confidence protein-protein interactions, we observed that all 8 cross-links involving residues in the nucleosome histone core can be mapped onto the high-resolution, atomic-level structure of the human nucleosome,⁶⁰ consistent with the 23 Å distance constraint based on the length of the DSSO linker (Figure 4C). Likewise, protein-protein interactions involving the core histone proteins captured many known histone-histone interactions, such as those between the core histone H2B and H2A, H4 and H3 histones (Figure 4C). Notably, we observed 46 cross-links involving the histone tails, which were not visualized in the isolated nucleosome core structure *in vitro*,⁶⁰ suggesting that they are in fact organized and bound to specific cofactors *in situ*. In particular, we observed numerous cross-links involving histone H2B K5 and histone H3 K4 and K18, which were also found in a recent study.⁶¹

Notably, using labeling of protein complexes *in situ*, we observed numerous previously unrecognized interactions. For example, we found numerous cross-links between the HMGN1 and HMGN2 proteins with histone H2B and histone H1.2 (Figure 5D & E). HMGN1 and HMGN2 are known to regulate chromatin configuration,⁶² which can involve phosphorylation of histone H2A and histone H3.⁶³⁻⁶⁴ Our findings indicate that HMGN1 can also bind directly to K120 of H2B, and HMGN2 can bind directly to K206 of H1.2 in the context of native chromatin *in situ*, which may provide the sought-after mechanism to explain its structural effects on

chromatin configuration directly.⁶⁵⁻⁶⁶ Likewise, we observed new protein-protein interactions involving ADARB1 with histone H2B, and centriolin/CNTRL with histone H3. This suggests that these two proteins may have unanticipated functions on DNA and/or chromatin. Similarly, we observed numerous cross-links to the linker histone H1, including PHRF1, ZNF786, PDE1C, and SEPT1, suggesting that these proteins may have additional structural chromatin functions, possibly in stabilizing or regulating histone linker-dependent higher-order chromatin conformations *in situ* (Figure 4E).

Lastly, we found numerous chromatin interactions of proteins involved in DNA damage repair and ribonucleoproteins. For instance, we observed cross-links from PARP1 to K108 and K116 of histone H2B, in addition to 17 intra-protein PARP1 cross-links (Figure 5A). We confirmed that 8 of 17 detected PARP1 cross-links could be mapped onto its high-resolution structure (PDB ID 4dqy),⁶⁷ as bound to a DNA double stranded break site (Figure 5B). PARP1 is known to self-associate,⁶⁸⁻⁶⁹ and our findings provide a direct mechanism of chromatin recruitment, linking PARP1 to histone H2B. This interaction may affect PARP1 self-association and DNA recognition and/or contribute to chromatin remodeling associated with DNA damage repair.⁷⁰⁻⁷⁴ Similarly, we found numerous cross-links among the heterogenous nuclear ribonucleoproteins hnRNPs and chromatin (Figure 5C). This included known interactions between hnRNPA1 and hnRNP2B, and between hnRNPA1 and hnRNPA3, providing specific structural details on their physical proximity.⁷⁵⁻⁷⁶ In addition, we observed direct interactions between DNA ligase LIG4 and hnRNPU proteins with histone H2B, raising the possibility that chromatin and hnRNP recruitment can regulate repair of DNA double-strand breaks.⁷⁷ While these interactions will require dedicated future studies to confirm and establish their functions, these results indicate

that native chromatin protein-protein interactions can be identified by using cross-linking mass spectrometry *in situ*.

Conclusions and Future Directions

Here, we sought to define the optimal methods for the isolation of cross-linked peptides from protein complexes labeled *in situ*, and to ensure their accurate identification using a target-decoy strategy adapted for cross-linking mass spectrometry. The presented generalized linear mixture model offers a facile means to control sensitivity and specificity based on a target-decoy spectral matching strategy adapted for cross-linking mass spectrometry. Using this approach, additional mixtures can be prepared in the future to model distinct protein complexes and peptide sequences, such as those containing particular chemical modifications. We also anticipate that improved scoring functions incorporating fragmentation features of specific cross-linkers as a function of fragmentation and isolation methods, such as the diagnostic sulfoxide fragment ion doublet in DSSO, as well as the use of isotopically labeled heavy and light cross-linkers to label experimental and control samples, can be used to improve the sensitivity and specificity of cross-linked peptide spectral identification.⁷⁸⁻⁸⁰ Likewise, for the study of specific complexes, sample-specific target databases can be used to improve spectral matching accuracy. Importantly, we found that the density and coverage of protein-protein interaction maps can be significantly improved through the use of diverse sequence-specific proteases, and complex cross-linked mixtures can be effectively analyzed using relatively fast CID-MS2/HCD-MS2 fragmentation methods. Lastly, our studies of native human chromatin labeled in live cells recapitulated known and revealed new protein interactions of nucleosomes and other chromatin-associated complexes. In all, this approach should facilitate the discovery and definition of protein-protein complexes *in vivo*, both in health and disease.

Acknowledgements

We thank Rosa Viner and David Horn for support of the beta version of XLinkX and Yael David for comments on the manuscript. This work was supported by the NIH R01 CA204396, P30 CA008748, Burroughs Wellcome Fund, Josie Robertson Investigator Program, Rita Allen Foundation, Alex's Lemonade Stand Foundation, St. Baldrick's Arceci Innovation Award, (A.K.), and Agency for Science, Technology and Research, Singapore (Z.S.). A.K. is the Damon Runyon-Richard Lumsden Foundation Clinical Investigator.

Notes

The authors declare no competing financial interest.

References

1. Kuzmanov, U.; Emili, A., Protein-protein interaction networks: probing disease mechanisms using model systems. *Genome Med* **2013**, *5* (4), 37.
2. Smits, A. H.; Vermeulen, M., Characterizing Protein-Protein Interactions Using Mass Spectrometry: Challenges and Opportunities. *Trends Biotechnol* **2016**, *34* (10), 825-34.
3. Huttlin, E. L.; Ting, L.; Bruckner, R. J.; Gebreab, F.; Gygi, M. P.; Szpyt, J.; Tam, S.; Zarraga, G.; Colby, G.; Baltier, K.; Dong, R.; Guarani, V.; Vaites, L. P.; Ordureau, A.; Rad, R.; Erickson, B. K.; Wuhr, M.; Chick, J.; Zhai, B.; Kolippakkam, D.; Mintseris, J.; Obar, R. A.; Harris, T.; Artavanis-Tsakonas, S.; Sowa, M. E.; De Camilli, P.; Paulo, J. A.; Harper, J. W.; Gygi, S. P., The BioPlex Network: A Systematic Exploration of the Human Interactome. *Cell* **2015**, *162* (2), 425-40.
4. Huttlin, E. L.; Bruckner, R. J.; Paulo, J. A.; Cannon, J. R.; Ting, L.; Baltier, K.; Colby, G.; Gebreab, F.; Gygi, M. P.; Parzen, H.; Szpyt, J.; Tam, S.; Zarraga, G.; Pontano-Vaites, L.; Swarup, S.; White, A. E.; Schweppe, D. K.; Rad, R.; Erickson, B. K.; Obar, R. A.; Guruharsha, K. G.; Li, K.; Artavanis-Tsakonas, S.; Gygi, S. P.; Harper, J. W., Architecture of the human interactome defines protein communities and disease networks. *Nature* **2017**, *545* (7655), 505-509.
5. Wan, C.; Borgeson, B.; Phanse, S.; Tu, F.; Drew, K.; Clark, G.; Xiong, X.; Kagan, O.; Kwan, J.; Bezginov, A.; Chessman, K.; Pal, S.; Cromar, G.; Papoulas, O.; Ni, Z.; Boutz, D. R.; Stoilova, S.; Havugimana, P. C.; Guo, X.; Maly, R. H.; Sarov, M.; Greenblatt, J.; Babu, M.; Derry, W. B.; Tillier, E. R.; Wallingford, J. B.; Parkinson, J.; Marcotte, E. M.; Emili, A., Panorama of ancient metazoan macromolecular complexes. *Nature* **2015**, *525* (7569), 339-44.
6. Havugimana, P. C.; Hart, G. T.; Nepusz, T.; Yang, H.; Turinsky, A. L.; Li, Z.; Wang, P. I.; Boutz, D. R.; Fong, V.; Phanse, S.; Babu, M.; Craig, S. A.; Hu, P.; Wan, C.; Vlasblom, J.; Dar, V. U.;

- Bezginov, A.; Clark, G. W.; Wu, G. C.; Wodak, S. J.; Tillier, E. R.; Paccanaro, A.; Marcotte, E. M.; Emili, A., A census of human soluble protein complexes. *Cell* **2012**, *150* (5), 1068-81.
7. Hein, M. Y.; Hubner, N. C.; Poser, I.; Cox, J.; Nagaraj, N.; Toyoda, Y.; Gak, I. A.; Weisswange, I.; Mansfeld, J.; Buchholz, F.; Hyman, A. A.; Mann, M., A human interactome in three quantitative dimensions organized by stoichiometries and abundances. *Cell* **2015**, *163* (3), 712-23.
8. 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* (3), 530-40.
9. Schneider, M.; Belsom, A.; Rappsilber, J., Protein Tertiary Structure by Crosslinking/Mass Spectrometry. *Trends Biochem Sci* **2018**, *43* (3), 157-169.
10. Muller, M. Q.; Dreiocker, F.; Ihling, C. H.; Schafer, M.; Sinz, A., Cleavable cross-linker for protein structure analysis: reliable identification of cross-linking products by tandem MS. *Anal Chem* **2010**, *82* (16), 6958-68.
11. Tang, X.; Bruce, J. E., A new cross-linking strategy: protein interaction reporter (PIR) technology for protein-protein interaction studies. *Mol Biosyst* **2010**, *6* (6), 939-47.
12. Kao, A.; Chiu, C. L.; Vellucci, D.; Yang, Y.; Patel, V. R.; Guan, S.; Randall, A.; Baldi, P.; Rychnovsky, S. D.; Huang, L., Development of a novel cross-linking strategy for fast and accurate identification of cross-linked peptides of protein complexes. *Mol Cell Proteomics* **2011**, *10* (1), M110 002212.
13. Kaake, R. M.; Wang, X.; Burke, A.; Yu, C.; Kandur, W.; Yang, Y.; Novtisky, E. J.; Second, T.; Duan, J.; Kao, A.; Guan, S.; Vellucci, D.; Rychnovsky, S. D.; Huang, L., A new in vivo cross-linking

mass spectrometry platform to define protein-protein interactions in living cells. *Mol Cell Proteomics* **2014**, *13* (12), 3533-43.

14. Kandur, W. V.; Kao, A.; Vellucci, D.; Huang, L.; Rychnovsky, S. D., Design of CID-cleavable protein cross-linkers: identical mass modifications for simpler sequence analysis. *Org Biomol Chem* **2015**, *13* (38), 9793-807.

15. Liu, F.; Rijkers, D. T.; Post, H.; Heck, A. J., Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry. *Nat Methods* **2015**, *12* (12), 1179-84.

16. Liu, F.; Lossel, P.; Scheltema, R.; Viner, R.; Heck, A. J. R., Optimized fragmentation schemes and data analysis strategies for proteome-wide cross-link identification. *Nat Commun* **2017**, *8*, 15473.

17. Gotze, M.; Pettelkau, J.; Fritzsche, R.; Ihling, C. H.; Schafer, M.; Sinz, A., Automated assignment of MS/MS cleavable cross-links in protein 3D-structure analysis. *J Am Soc Mass Spectrom* **2015**, *26* (1), 83-97.

18. Arlt, C.; Gotze, M.; Ihling, C. H.; Hage, C.; Schafer, M.; Sinz, A., Integrated Workflow for Structural Proteomics Studies Based on Cross-Linking/Mass Spectrometry with an MS/MS Cleavable Cross-Linker. *Anal Chem* **2016**, *88* (16), 7930-7.

19. Mohr, J. P.; Perumalla, P.; Chavez, J. D.; Eng, J. K.; Bruce, J. E., Mango: A General Tool for Collision Induced Dissociation-Cleavable Cross-Linked Peptide Identification. *Anal Chem* **2018**, *90* (10), 6028-6034.

20. Yang, B.; Wu, Y. J.; Zhu, M.; Fan, S. B.; Lin, J.; Zhang, K.; Li, S.; Chi, H.; Li, Y. X.; Chen, H. F.; Luo, S. K.; Ding, Y. H.; Wang, L. H.; Hao, Z.; Xiu, L. Y.; Chen, S.; Ye, K.; He, S. M.; Dong, M. Q., Identification of cross-linked peptides from complex samples. *Nat Methods* **2012**, *9* (9), 904-6.

21. Tan, D.; Li, Q.; Zhang, M. J.; Liu, C.; Ma, C.; Zhang, P.; Ding, Y. H.; Fan, S. B.; Tao, L.; Yang, B.; Li, X.; Ma, S.; Liu, J.; Feng, B.; Liu, X.; Wang, H. W.; He, S. M.; Gao, N.; Ye, K.; Dong, M. Q.; Lei, X., Trifunctional cross-linker for mapping protein-protein interaction networks and comparing protein conformational states. *Elife* **2016**, *5*.
22. Zheng, C.; Yang, L.; Hoopmann, M. R.; Eng, J. K.; Tang, X.; Weisbrod, C. R.; Bruce, J. E., Cross-linking measurements of in vivo protein complex topologies. *Mol Cell Proteomics* **2011**, *10* (10), M110 006841.
23. Zhong, X.; Navare, A. T.; Chavez, J. D.; Eng, J. K.; Schweppe, D. K.; Bruce, J. E., Large-Scale and Targeted Quantitative Cross-Linking MS Using Isotope-Labeled Protein Interaction Reporter (PIR) Cross-Linkers. *J Proteome Res* **2017**, *16* (2), 720-727.
24. Chavez, J. D.; Weisbrod, C. R.; Zheng, C.; Eng, J. K.; Bruce, J. E., Protein interactions, post-translational modifications and topologies in human cells. *Mol Cell Proteomics* **2013**, *12* (5), 1451-67.
25. Liu, F.; Lossel, P.; Rabbitts, B. M.; Balaban, R. S.; Heck, A. J. R., The interactome of intact mitochondria by cross-linking mass spectrometry provides evidence for coexisting respiratory supercomplexes. *Mol Cell Proteomics* **2018**, *17* (2), 216-232.
26. Wu, X.; Chavez, J. D.; Schweppe, D. K.; Zheng, C.; Weisbrod, C. R.; Eng, J. K.; Murali, A.; Lee, S. A.; Ramage, E.; Gallagher, L. A.; Kulasekara, H. D.; Edrozo, M. E.; Kamischke, C. N.; Brittnacher, M. J.; Miller, S. I.; Singh, P. K.; Manoil, C.; Bruce, J. E., In vivo protein interaction network analysis reveals porin-localized antibiotic inactivation in *Acinetobacter baumannii* strain AB5075. *Nat Commun* **2016**, *7*, 13414.

27. Wisniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M., Universal sample preparation method for proteome analysis. *Nat Methods* **2009**, *6* (5), 359-62.
28. Wisniewski, J. R.; Zielinska, D. F.; Mann, M., Comparison of ultrafiltration units for proteomic and N-glycoproteomic analysis by the filter-aided sample preparation method. *Anal Biochem* **2011**, *410* (2), 307-9.
29. Dhabaria, A.; Cifani, P.; Reed, C.; Steen, H.; Kentsis, A., A High-Efficiency Cellular Extraction System for Biological Proteomics. *J Proteome Res* **2015**, *14* (8), 3403-8.
30. Dhabaria, A., Cifani, P., Kentsis, A., Fabrication of Capillary Columns with Integrated Frits for Mass Spectrometry. Protocol Exchange, 2015.
31. Cifani, P., Dhabaria, A., Kentsis, A., Fabrication of Nanoelectrospray Emitters for LC-MS. 2015.
32. The UniProt, C., UniProt: the universal protein knowledgebase. *Nucleic Acids Res* **2017**, *45* (D1), D158-D169.
33. Mellacheruvu, D.; Wright, Z.; Couzens, A. L.; Lambert, J. P.; St-Denis, N. A.; Li, T.; Miteva, Y. V.; Hauri, S.; Sardi, M. E.; Low, T. Y.; Halim, V. A.; Bagshaw, R. D.; Hubner, N. C.; Al-Hakim, A.; Bouchard, A.; Faubert, D.; Fermin, D.; Dunham, W. H.; Goudreault, M.; Lin, Z. Y.; Badillo, B. G.; Pawson, T.; Durocher, D.; Coulombe, B.; Aebersold, R.; Superti-Furga, G.; Colinge, J.; Heck, A. J.; Choi, H.; Gstaiger, M.; Mohammed, S.; Cristea, I. M.; Bennett, K. L.; Washburn, M. P.; Raught, B.; Ewing, R. M.; Gingras, A. C.; Nesvizhskii, A. I., The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. *Nat Methods* **2013**, *10* (8), 730-6.

34. Kall, L.; Canterbury, J. D.; Weston, J.; Noble, W. S.; MacCoss, M. J., Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat Methods* **2007**, *4* (11), 923-5.
35. Oliveros, J. C., Venny. An interactive tool for comparing lists with Venn's diagrams. In <http://bioinfogp.cnb.csic.es/tools/venny/index.html>, 2007-2015.
36. Combe, C. W.; Fischer, L.; Rappsilber, J., xiNET: cross-link network maps with residue resolution. *Mol Cell Proteomics* **2015**, *14* (4), 1137-47.
37. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **2004**, *25* (13), 1605-12.
38. Kosinski, J.; von Appen, A.; Ori, A.; Karius, K.; Muller, C. W.; Beck, M., Xlink Analyzer: software for analysis and visualization of cross-linking data in the context of three-dimensional structures. *J Struct Biol* **2015**, *189* (3), 177-83.
39. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T., Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **2003**, *13* (11), 2498-504.
40. Bhattacharya, M.; Jain, N.; Mukhopadhyay, S., Insights into the mechanism of aggregation and fibril formation from bovine serum albumin. *J Phys Chem B* **2011**, *115* (14), 4195-205.
41. Vetri, V.; Librizzi, F.; Leone, M.; Militello, V., Thermal aggregation of bovine serum albumin at different pH: comparison with human serum albumin. *Eur Biophys J* **2007**, *36* (7), 717-25.

42. Bujacz, A., Structures of bovine, equine and leporine serum albumin. *Acta Crystallogr D Biol Crystallogr* **2012**, *68* (Pt 10), 1278-89.
43. Shen, Y.; Tolic, N.; Xie, F.; Zhao, R.; Purvine, S. O.; Schepmoes, A. A.; Moore, R. J.; Anderson, G. A.; Smith, R. D., Effectiveness of CID, HCD, and ETD with FT MS/MS for degradomic-peptidomic analysis: comparison of peptide identification methods. *J Proteome Res* **2011**, *10* (9), 3929-43.
44. Good, D. M.; Wirtala, M.; McAlister, G. C.; Coon, J. J., Performance characteristics of electron transfer dissociation mass spectrometry. *Mol Cell Proteomics* **2007**, *6* (11), 1942-51.
45. Kolbowski, L.; Mendes, M. L.; Rappsilber, J., Optimizing the Parameters Governing the Fragmentation of Cross-Linked Peptides in a Tribid Mass Spectrometer. *Anal Chem* **2017**, *89* (10), 5311-5318.
46. Sticker, A.; Martens, L.; Clement, L., Mass spectrometrists should search for all peptides, but assess only the ones they care about. *Nat Methods* **2017**, *14* (7), 643-644.
47. Noble, W. S., Mass spectrometrists should search only for peptides they care about. *Nat Methods* **2015**, *12* (7), 605-8.
48. Noble, W. S.; Keich, U., Response to "Mass spectrometrists should search for all peptides, but assess only the ones they care about". *Nat Methods* **2017**, *14* (7), 644.
49. Knudsen, G. M.; Chalkley, R. J., The effect of using an inappropriate protein database for proteomic data analysis. *PLoS One* **2011**, *6* (6), e20873.
50. Giansanti, P.; Tsiatsiani, L.; Low, T. Y.; Heck, A. J., Six alternative proteases for mass spectrometry-based proteomics beyond trypsin. *Nat Protoc* **2016**, *11* (5), 993-1006.
51. Tsiatsiani, L.; Heck, A. J., Proteomics beyond trypsin. *FEBS J* **2015**, *282* (14), 2612-26.

52. Swaney, D. L.; Wenger, C. D.; Coon, J. J., Value of using multiple proteases for large-scale mass spectrometry-based proteomics. *J Proteome Res* **2010**, *9* (3), 1323-9.
53. Kornberg, R. D.; LaPointe, J. W.; Lorch, Y., Preparation of nucleosomes and chromatin. *Methods Enzymol* **1989**, *170*, 3-14.
54. Hewish, D. R.; Burgoyne, L. A., Chromatin sub-structure. The digestion of chromatin DNA at regularly spaced sites by a nuclear deoxyribonuclease. *Biochem Biophys Res Commun* **1973**, *52* (2), 504-10.
55. Zaret, K., Micrococcal nuclease analysis of chromatin structure. *Curr Protoc Mol Biol* **2005**, *Chapter 21*, Unit 21 1.
56. Stark, C.; Breitkreutz, B. J.; Reguly, T.; Boucher, L.; Breitkreutz, A.; Tyers, M., BioGRID: a general repository for interaction datasets. *Nucleic Acids Res* **2006**, *34* (Database issue), D535-9.
57. Chatr-Aryamontri, A.; Oughtred, R.; Boucher, L.; Rust, J.; Chang, C.; Kolas, N. K.; O'Donnell, L.; Oster, S.; Theesfeld, C.; Sellam, A.; Stark, C.; Breitkreutz, B. J.; Dolinski, K.; Tyers, M., The BioGRID interaction database: 2017 update. *Nucleic Acids Res* **2017**, *45* (D1), D369-D379.
58. Ruepp, A.; Brauner, B.; Dunger-Kaltenbach, I.; Frishman, G.; Montrone, C.; Stransky, M.; Waegele, B.; Schmidt, T.; Doudieu, O. N.; Stumpflen, V.; Mewes, H. W., CORUM: the comprehensive resource of mammalian protein complexes. *Nucleic Acids Res* **2008**, *36* (Database issue), D646-50.
59. Ruepp, A.; Waegele, B.; Lechner, M.; Brauner, B.; Dunger-Kaltenbach, I.; Fobo, G.; Frishman, G.; Montrone, C.; Mewes, H. W., CORUM: the comprehensive resource of mammalian protein complexes--2009. *Nucleic Acids Res* **2010**, *38* (Database issue), D497-501.

60. Tachiwana, H.; Osakabe, A.; Shiga, T.; Miya, Y.; Kimura, H.; Kagawa, W.; Kurumizaka, H., Structures of human nucleosomes containing major histone H3 variants. *Acta Crystallogr D Biol Crystallogr* **2011**, *67* (Pt 6), 578-83.
61. Fasci, D.; van Ingen, H.; Scheltema, R. A.; Heck, A. J. R., Histone interaction landscapes visualized by crosslinking mass spectrometry in intact cell nuclei. *Mol Cell Proteomics* **2018**.
62. Murphy, K. J.; Cutter, A. R.; Fang, H.; Postnikov, Y. V.; Bustin, M.; Hayes, J. J., HMGN1 and 2 remodel core and linker histone tail domains within chromatin. *Nucleic Acids Res* **2017**, *45* (17), 9917-9930.
63. Lim, J. H.; Catez, F.; Birger, Y.; West, K. L.; Prymakowska-Bosak, M.; Postnikov, Y. V.; Bustin, M., Chromosomal protein HMGN1 modulates histone H3 phosphorylation. *Mol Cell* **2004**, *15* (4), 573-84.
64. Postnikov, Y. V.; Belova, G. I.; Lim, J. H.; Bustin, M., Chromosomal protein HMGN1 modulates the phosphorylation of serine 1 in histone H2A. *Biochemistry* **2006**, *45* (50), 15092-9.
65. Shimahara, H.; Hirano, T.; Ohya, K.; Matsuta, S.; Seeram, S. S.; Tate, S., Nucleosome structural changes induced by binding of non-histone chromosomal proteins HMGN1 and HMGN2. *FEBS Open Bio* **2013**, *3*, 184-91.
66. Postnikov, Y. V.; Bustin, M., Functional interplay between histone H1 and HMG proteins in chromatin. *Biochim Biophys Acta* **2016**, *1859* (3), 462-7.
67. Langelier, M. F.; Planck, J. L.; Roy, S.; Pascal, J. M., Structural basis for DNA damage-dependent poly(ADP-ribosyl)ation by human PARP-1. *Science* **2012**, *336* (6082), 728-32.

68. Wacker, D. A.; Ruhl, D. D.; Balagamwala, E. H.; Hope, K. M.; Zhang, T.; Kraus, W. L., The DNA binding and catalytic domains of poly(ADP-ribose) polymerase 1 cooperate in the regulation of chromatin structure and transcription. *Mol Cell Biol* **2007**, *27* (21), 7475-85.
69. Bonfiglio, J. J.; Fontana, P.; Zhang, Q.; Colby, T.; Gibbs-Seymour, I.; Atanassov, I.; Bartlett, E.; Zaja, R.; Ahel, I.; Matic, I., Serine ADP-Ribosylation Depends on HPF1. *Mol Cell* **2017**, *65* (5), 932-940 e6.
70. Poirier, G. G.; de Murcia, G.; Jongstra-Bilen, J.; Niedergang, C.; Mandel, P., Poly(ADP-ribose)ylation of polynucleosomes causes relaxation of chromatin structure. *Proc Natl Acad Sci U S A* **1982**, *79* (11), 3423-7.
71. D'Amours, D.; Desnoyers, S.; D'Silva, I.; Poirier, G. G., Poly(ADP-ribose)ylation reactions in the regulation of nuclear functions. *Biochem J* **1999**, *342* (Pt 2), 249-68.
72. Messner, S.; Altmeyer, M.; Zhao, H.; Pozivil, A.; Roschitzki, B.; Gehrig, P.; Rutishauser, D.; Huang, D.; Caflisch, A.; Hottiger, M. O., PARP1 ADP-ribosylates lysine residues of the core histone tails. *Nucleic Acids Res* **2010**, *38* (19), 6350-62.
73. Huambachano, O.; Herrera, F.; Rancourt, A.; Satoh, M. S., Double-stranded DNA binding domain of poly(ADP-ribose) polymerase-1 and molecular insight into the regulation of its activity. *J Biol Chem* **2011**, *286* (9), 7149-60.
74. Ray Chaudhuri, A.; Nussenzweig, A., The multifaceted roles of PARP1 in DNA repair and chromatin remodelling. *Nat Rev Mol Cell Biol* **2017**, *18* (10), 610-621.
75. Roy, R.; Durie, D.; Li, H.; Liu, B. Q.; Skehel, J. M.; Mauri, F.; Cuorvo, L. V.; Barbareschi, M.; Guo, L.; Holcik, M.; Seckl, M. J.; Pardo, O. E., hnRNPA1 couples nuclear export and translation of specific mRNAs downstream of FGF-2/S6K2 signalling. *Nucleic Acids Res* **2014**, *42* (20), 12483-97.

76. Jurica, M. S.; Licklider, L. J.; Gygi, S. R.; Grigorieff, N.; Moore, M. J., Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. *RNA* **2002**, *8* (4), 426-39.
77. Polo, S. E.; Blackford, A. N.; Chapman, J. R.; Baskcomb, L.; Gravel, S.; Rusch, A.; Thomas, A.; Blundred, R.; Smith, P.; Kzhyshkowska, J.; Dobner, T.; Taylor, A. M.; Turnell, A. S.; Stewart, G. S.; Grand, R. J.; Jackson, S. P., Regulation of DNA-end resection by hnRNPU-like proteins promotes DNA double-strand break signaling and repair. *Mol Cell* **2012**, *45* (4), 505-16.
78. Ivanov, M. V.; Levitsky, L. I.; Lobas, A. A.; Panic, T.; Laskay, U. A.; Mitulovic, G.; Schmid, R.; Pridatchenko, M. L.; Tsybin, Y. O.; Gorshkov, M. V., Empirical multidimensional space for scoring peptide spectrum matches in shotgun proteomics. *J Proteome Res* **2014**, *13* (4), 1911-20.
79. Frank, A. M., A ranking-based scoring function for peptide-spectrum matches. *J Proteome Res* **2009**, *8* (5), 2241-52.
80. Fischer, L.; Rappsilber, J., Quirks of Error Estimation in Cross-Linking/Mass Spectrometry. *Anal Chem* **2017**, *89* (7), 3829-3833.

Figures

Figure 1

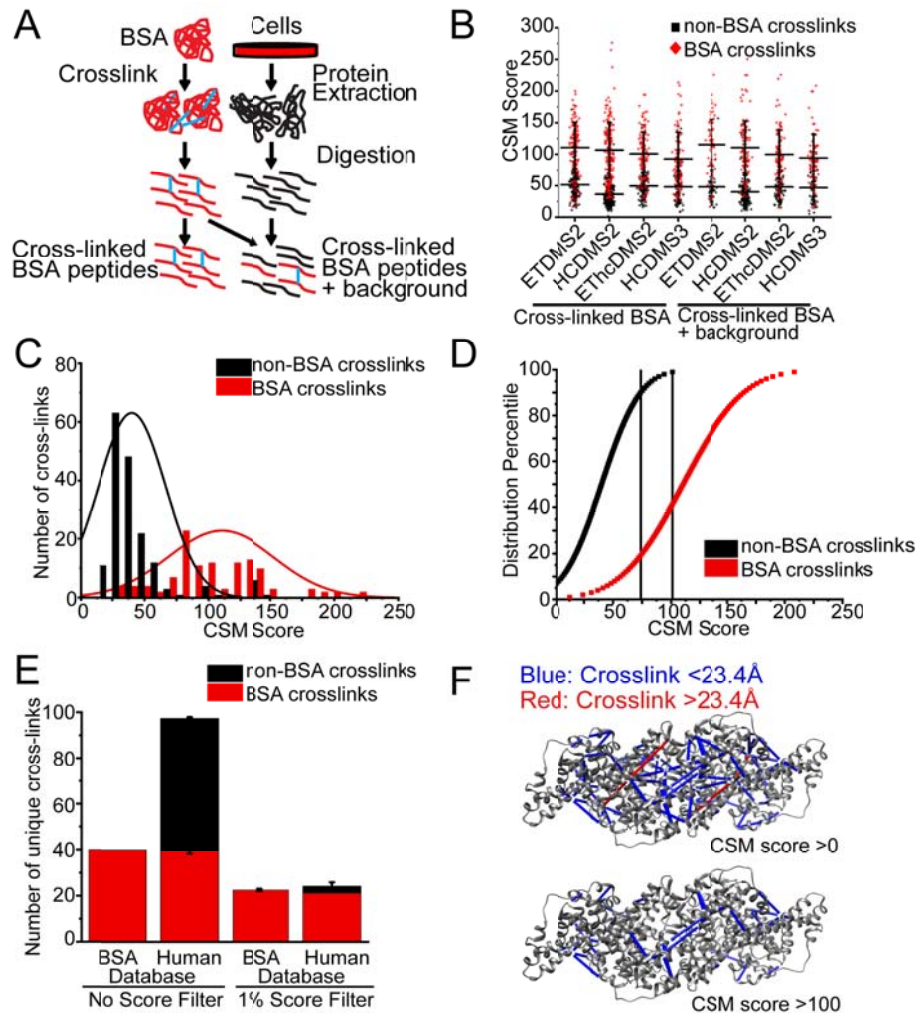


Figure 1. Calibration of cross-link peptide-spectra matching (CSM) scores to control sensitivity and specificity of large-scale XL-MS. A) Workflow for preparation of cross-linked BSA peptides and cross-linked BSA peptides spiked into non-cross-linked proteome background peptides. B) High number of BSA cross-linked peptides (red) and non-BSA cross-linked peptides (black) identified across fragmentation methods and in both cross-linked BSA and cross-linked BSA spiked into background. C) Fitted Gaussian distribution of cross-link spectra matching score of non-BSA cross-links and BSA cross-links for cross-linked BSA only sample acquired with CID-MS2/HCD-MS2 fragmentation. D) Percentile plot of gaussian distribution of CSM score with 1% and 10% score filter to eliminate 99% and 90% of non-BSA cross-links. E) Comparison of number of cross-links identified when no score filter is applied and when 1% score filter is applied F) Cross-links violating physical distance constraint of cross-linker (red) are eliminated when 1% score filter is applied when mapping cross-links to crystal structure (PDB ID 4f5s).

Figure 2

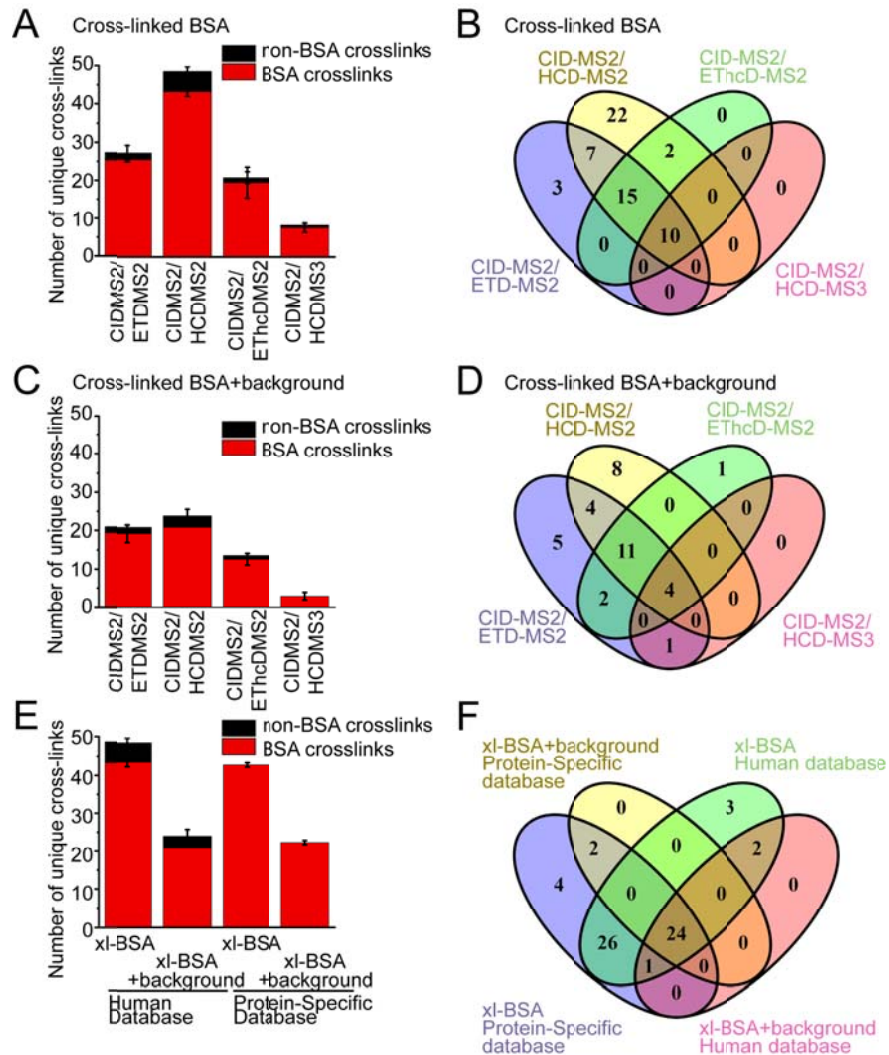


Figure 2. Improved cross-linked peptide identification using CID-MS2/HCD-MS2 fragmentation. Comparison of 4 different fragmentation methods: CID-MS2/HCD-MS2, CID-MS2/ETD-MS2, CID-MS2/ETThcD-MS2 and CID-MS2/HCD-MS3 and databases: BSA protein specific database or human proteome database with BSA for increasing cross-link peptide identification. A) Comparison between 4 fragmentation methods for average number of cross-link identifications of cross-linked BSA sample. BSA cross-links colored red and non-BSA cross-links colored black. Error bars represent standard deviation from three biological replicates. B) Comparison between 4 fragmentation methods for overlap of all unique cross-link identifications C) Comparison between 4 fragmentation methods for average number of cross-link identifications of cross-linked BSA peptides spiked into human proteome background. D) Comparison between 4 fragmentation methods for overlap of all unique cross-link identifications of cross-linked BSA sample spiked into human proteome background. E) Comparison between 2 databases for average number of cross-link identifications of cross-linked BSA peptides and cross-linked BSA peptides spiked into human proteome background. F) Comparison between 2 databases for overlap of all unique cross-link identifications of cross-linked BSA peptides and cross-linked BSA peptides spiked into human proteome background.

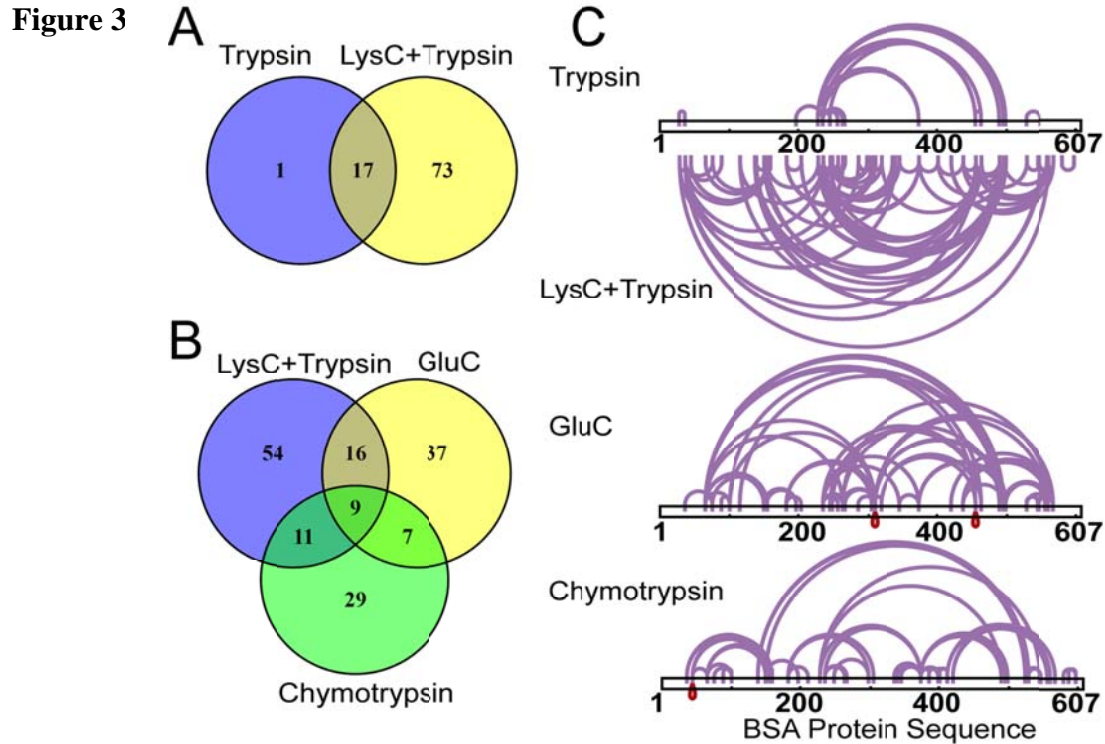


Figure 3. Multiple proteases to expand the coverage and density of DSSO XL-MS protein interaction maps. A) Combined LysC and trypsin protease digestion greatly increases number of cross-link peptides identified compared to trypsin only protease digestion, B) Low number of overlapping cross-linked peptides identified from LysC and trypsin or GluC only or chymotrypsin only protease digestion, C) Visualization of BSA cross-link sites identified from trypsin only, LysC and trypsin, GluC only or chymotrypsin only protease digestion, as analyzed using CID-MS2/HCD-MS2 fragmentation.

Figure 4

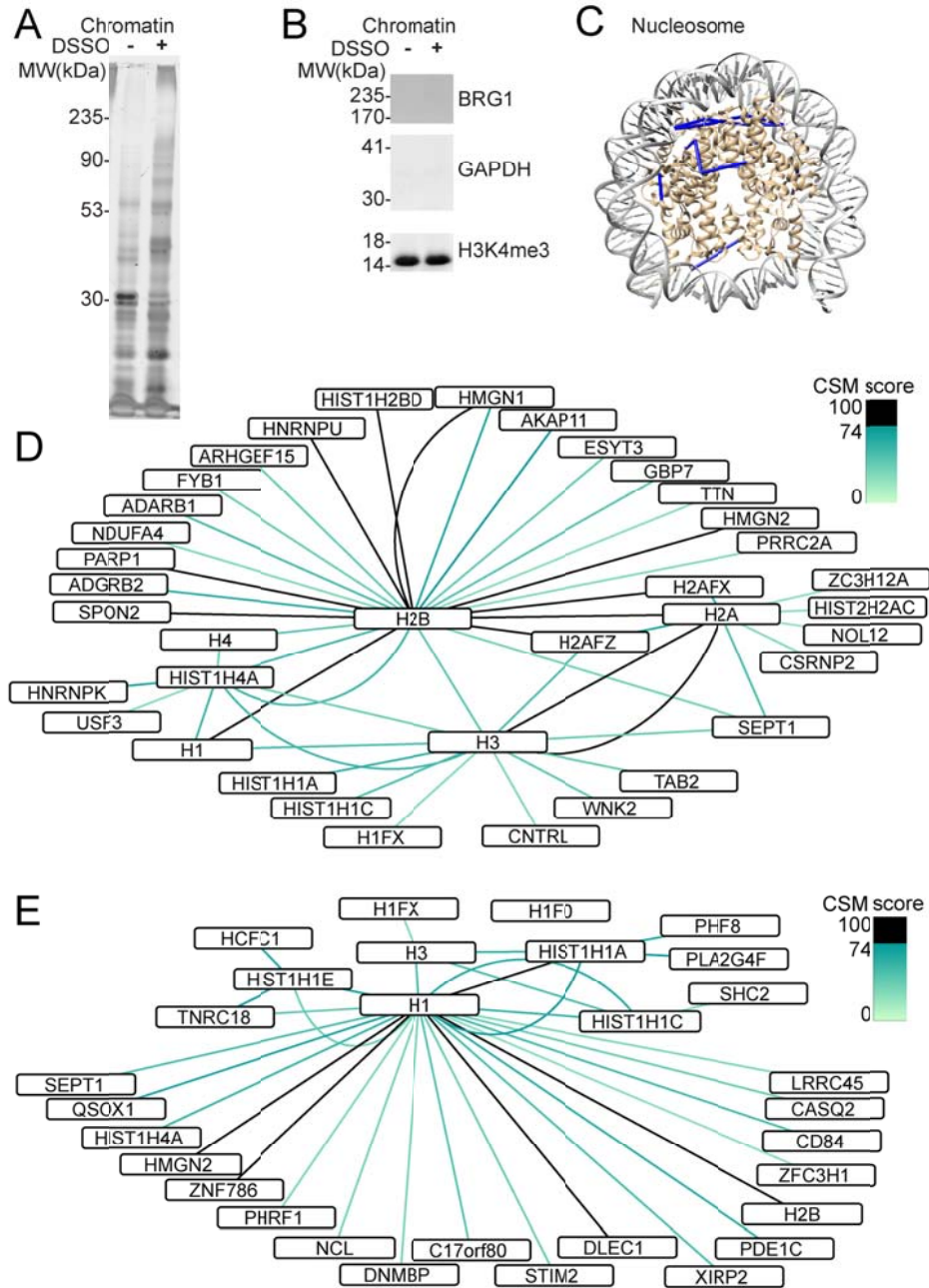


Figure 4. Native chromatin protein-protein interactions identified using cross-linking mass spectrometry *in situ*. A) Purified chromatin bound proteins are cross-linked by cross-linking *in situ*. B) Enrichment of histone protein (H3K4me3) with absence of nuclear protein (BRG1) and cytoplasmic protein (GAPDH) from chromatin bound proteins by western blot. C) Cross-links identified map onto known high resolution nucleosome structure (PDB ID 3av1). D) Map of protein-protein interactions between core histone proteins (H2A, H2B, H3, H4) colored according to CSM score. E) Map of protein-protein interactions involving linker histone H1.

Figure 5

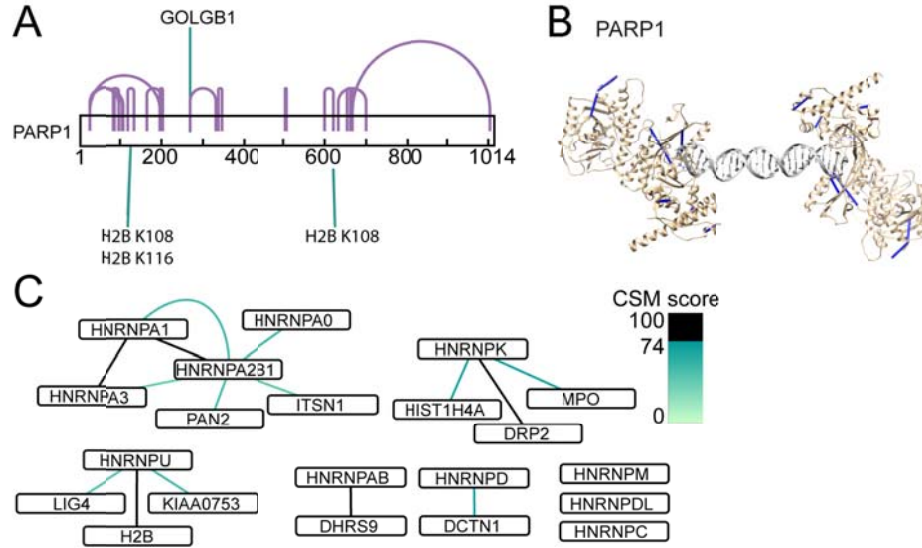


Figure 5. Chromatin interactions of proteins involved in DNA damage repair and ribonucleoproteins. A) DNA repair protein PARP1 interacts with Histone H2B. B) PARP1 cross-links map onto known high resolution structure of PARP1 bound to DNA double stranded break (PDB ID 4dqy). C) Protein-protein interactions of RNA-binding heterogeneous nuclear ribonucleoproteins (hnRNPs).

Figure Legends

Figure 1. Calibration of cross-link peptide-spectra matching (CSM) scores to control sensitivity and specificity of large-scale XL-MS. A) Workflow for preparation of cross-linked BSA peptides and cross-linked BSA peptides spiked into non-cross-linked proteome background peptides. B) High number of BSA cross-linked peptides (red) and non-BSA cross-linked peptides (black) identified across fragmentation methods and in both cross-linked BSA and cross-linked BSA spiked into background. C) Fitted Gaussian distribution of cross-link spectra matching score of non-BSA cross-links and BSA cross-links for cross-linked BSA only sample acquired with CID-MS2/HCD-MS2 fragmentation. D) Percentile plot of gaussian distribution of CSM score with 1% and 10% score filter to eliminate 99% and 90% of non-BSA cross-links. E) Comparison of number of cross-links identified when no score filter is applied and when 1% score filter is applied F) Cross-links violating physical distance constraint of cross-linker (red) are eliminated when 1% score filter is applied when mapping cross-links to crystal structure (PDB ID 4f5s).

Figure 2. Improved cross-linked peptide identification using CID-MS2/HCD-MS2 fragmentation. Comparison of 4 different fragmentation methods: CID-MS2/HCD-MS2, CID-MS2/ETD-MS2, CID-MS2/EThcD-MS2 and CID-MS2/HCD-MS3 and databases: BSA protein specific database or human proteome database with BSA for increasing cross-link peptide identification. A) Comparison between 4 fragmentation methods for average number of cross-link identifications of cross-linked BSA sample. BSA cross-links colored red and non-BSA cross-links colored black. Error bars represent standard deviation from three biological replicates. B) Comparison between 4 fragmentation methods for overlap of all unique cross-link identifications C) Comparison between 4 fragmentation methods for average number of cross-

link identifications of cross-linked BSA peptides spiked into human proteome background. D) Comparison between 4 fragmentation methods for overlap of all unique cross-link identifications of cross-linked BSA sample spiked into human proteome background. E) Comparison between 2 databases for average number of cross-link identifications of cross-linked BSA peptides and cross-linked BSA peptides spiked into human proteome background. F) Comparison between 2 databases for overlap of all unique cross-link identifications of cross-linked BSA peptides and cross-linked BSA peptides spiked into human proteome background.

Figure 3. Multiple proteases to expand the coverage and density of DSSO XL-MS protein interaction maps. A) Combined LysC and trypsin protease digestion greatly increases number of cross-link peptides identified compared to trypsin only protease digestion, B) Low number of overlapping cross-linked peptides identified from LysC and trypsin or GluC only or chymotrypsin only protease digestion, C) Visualization of BSA cross-link sites identified from trypsin only, LysC and trypsin, GluC only or chymotrypsin only protease digestion, as analyzed using CID-MS2/HCD-MS2 fragmentation.

Figure 4. Native chromatin protein-protein interactions identified using cross-linking mass spectrometry *in situ*. A) Purified chromatin bound proteins are cross-linked by cross-linking *in situ*. B) Enrichment of histone protein (H3K4me3) with absence of nuclear protein (BRG1) and cytoplasmic protein (GAPDH) from chromatin bound proteins by western blot. C) Cross-links identified map onto known high resolution nucleosome structure (PDB ID 3av1). D) Map of protein-protein interactions between core histone proteins (H2A, H2B, H3, H4) colored according to CSM score. E) Map of protein-protein interactions involving linker histone H1.

Figure 5. Chromatin interactions of proteins involved in DNA damage repair and ribonucleoproteins. A) DNA repair protein PARP1 interacts with Histone H2B. B) PARP1 cross-links map onto known high resolution structure of PARP1 bound to DNA double stranded break (PDB ID 4dqy). C) Protein-protein interactions of RNA-binding heterogeneous nuclear ribonucleoproteins (hnRNPs).

Supporting Information

Figure S-1. Characterization of peptide length, m/z and charge state of cross-linked peptides from comparison of fragmentation methods: CID-MS2/HCD-MS2, CID-MS2/ETD-MS2, CID-MS2/EThcD-MS2 and CID-MS2/HCD-MS3. A) Cross-link peptide length by number of amino acids from cross-linked BSA peptide spectra. B) Precursor m/z values. C) Number of cross-linked peptide spectra based on charge state. D,E,F,G) Comparison of peptide A and peptide B length from cross-linked peptides identified by CID-MS2/HCD-MS2, CID-MS2/ETD-MS2, CID-MS2/EThcD-MS2 or CID-MS2/HCD-MS3.

Figure S-2. Characterization of peptide length, m/z and charge state of cross-linked peptides from comparison of multiple protease digestion: Trypsin only, combined LysC and Trypsin, chymotrypsin only, GluC only, as analyzed using CID-MS2/HCD-MS2 fragmentation. A) Cross-link peptide length by number of amino acids from cross-linked BSA peptide spectra. B) Precursor m/z value. C) Number of cross-linked peptide spectra based on charge state. D,E,F,G) Comparison of peptide A and peptide B length from cross-linked peptides identified by trypsin only, combined LysC and trypsin, chymotrypsin only or GluC only protease digestion.

Table S-1. List of protein cross-links from in-situ cross-linking of purified chromatin identified from at least 2 out of 3 biological replicates.