

Protein Interactions, Post-translational Modifications and Topologies in Human Cells*[§]

Juan D. Chavez‡, Chad R. Weisbrod‡, Chunxiang Zheng‡, Jimmy K. Eng‡, and James E. Bruce‡§

The unique and remarkable physicochemical properties of protein surface topologies give rise to highly specific biomolecular interactions, which form the framework through which living systems are able to carry out their vast array of functions. Technological limitations undermine efforts to probe protein structures and interactions within unperturbed living systems on a large scale. Rapid chemical stabilization of proteins and protein complexes through chemical cross-linking offers the alluring possibility to study details of the protein structure to function relationships as they exist within living cells. Here we apply the latest technological advances in chemical cross-linking combined with mass spectrometry to study protein topologies and interactions from living human cells identifying a total of 368 cross-links. These include cross-links from all major cellular compartments including membrane, cytosolic and nuclear proteins. Intraprotein and interprotein cross-links were also observed for core histone proteins, including several cross-links containing post-translational modifications which are known histone marks conferring distinct epigenetic functions. Excitingly, these results demonstrate the applicability of cross-linking to make direct topological measurements on post-translationally modified proteins. The results presented here provide new details on the structures of known multi-protein complexes as well as evidence for new protein-protein interactions. *Molecular & Cellular Proteomics* 12: 10.1074/mcp.M112.024497, 1451–1467, 2013.

Proteins are the principal operatives within cells, involved in carrying out essentially all biological functions. A complex network of intra- and intermolecular interactions, post-translational modifications and abundance levels is required to maintain the delicate balance of function essential for life. Subtle changes within this network can give rise to specific biological responses to environmental factors, onset of disease, normal aging, and other biological processes. Therefore, direct experimental observation of protein structures and

interactions in relation to biological function is paramount to improved understanding of living systems.

Chemical cross-linking has long been used as a method of fixation to preserve biological samples in the fields of histology and pathology (1). Protein interactions and topologies have also been studied with chemical cross-linking methods for many years (2–4). Chemical cross-linking with mass spectrometry (XL-MS)¹ is emerging as a powerful technology to study protein structures and interactions in complex biological systems (5). Technological advances in chemistry, analytical instrumentation, and informatics are beginning to allow the successful application of XL-MS to study protein topologies and interactions on a large scale in complex biological systems. These methods are able to provide low resolution spatial information on protein topologies through distance constraints imposed by the chemical linker arm distance. The resultant distance constraints are often used to refine crystal structure measurements and to assist *de novo* structure prediction with molecular modeling techniques (6, 7). Structural information derived through cross-linking experiments is largely complementary to structural information obtained through other techniques including hydrogen-deuterium exchange mass spectrometry, NMR, and x-ray crystallography (8). In addition to structural information, cross-linking provides information about interacting proteins that is complimentary to information from techniques such as yeast two hybrid (9) and co-immunoprecipitation (co-IP)(10). One application of XL-MS that has seen recent success is the isolation of protein complexes from cells followed by *in vitro* cross-linking to study the protein complex architecture as demonstrated on the *E. coli* ribosome and human phosphatase 2A (11, 12). These applications illustrate the powerful utility XL-MS methods hold for acquiring topological data on complexes that are not completely amenable to interrogation via conventional structural biology techniques. Cross-linking was also used to provide the first structural insight into intact infectious virion

From the ‡Department of Genome Science, University of Washington, Seattle, Washington 98109

Received October 3, 2012, and in revised form, December 12, 2012

Published, MCP Papers in Press, January 25, 2013, DOI 10.1074/mcp.M112.024497

¹ The abbreviations used are: PIR, protein interaction reporter; LC-MS, liquid chromatography mass spectrometry; XL-MS, cross-linking with mass spectrometry; PDB, Protein Data Bank; ReACT, Real-time Analysis for Cross-linking Technology; co-IP, co-immunoprecipitation; NMR, Nuclear Magnetic Resonance; BDP, biotinylated aspartyl prolyl cross-linker; Fmoc, fluorenylmethyloxycarbonyl; NHP, N-hydroxyphthalimide; RP, resolving power.

particles of the *Potato Leafroll Virus* elucidating key topological features for virus assembly and transmission (13).

The extension of cross-linking technology to enable *in vivo* measurements is an attractive option to study protein topologies as they exist in their native environment, particularly for proteins with disordered domains or membrane proteins that may not maintain native folding states during purification. The use of XL-MS to study protein structures and interactions directly from living cells was first demonstrated on the bacterial system *Shewanella oneidensis*, where new details on the membrane protein electron transport machinery used during anaerobic respiration were discovered (14). More recently XL-MS has been applied to *E. coli* cell lysates (15, 16) as well as intact *E. coli* cells (17, 18). Importantly, these efforts have provided valuable information on protein topologies and complexes as they exist *in vivo* in contrast to studies on purified forms of proteins where endogenous topological features may not be preserved.

Recently, the application of cross-linking to complex systems has been further advanced by new mass spectrometric methods such as dynamic targeting of cross-linked species in real-time during the data acquisition (19). This new method has expanded the range of XL-MS experiments resulting in the identification of hundreds of cross-linked peptide pairs from cross-linked *E. coli* cells. The large-scale data sets that emerge from such experiments have necessitated the development of new software tools to aid in data analysis. XLink-DB is an online database repository that includes protein interaction network and protein structural analysis tools specifically designed for large scale cross-linking data sets (20). Tools such as XLink-DB make possible analyses that were prohibitively time consuming to perform manually such as mapping large numbers of cross-links onto available crystal structures in the Protein Data Bank (PDB). The development of chemical cross-linking for applications in measuring *in vivo* protein interaction topologies has been highlighted in a recent review (21).

In the presented work, the goal was to investigate the potential of XL-MS to more complex biological systems, in particular living human cells. Mammalian systems pose tremendous challenges to XL-MS because of the sheer size of the proteome and added complexity of post-translational modifications. Application of XL-MS techniques to complex systems offers the alluring possibility to probe the human disease interactome (22), while maintaining the ability to investigate post-translational modifications and their effects on *in vivo* interaction topologies. Here we report an initial mammalian (*H. sapiens*) protein interaction network inferred exclusively using mass spectrometry data from cross-linking experiments. Several cross-linked sites identified were found to carry specific post-translational modifications (PTM), including acetylated and methylated peptides from core histone proteins. Of the 115 identified cross-linked sites obtained from histone proteins 56 were found to contain one or more

PTM. Post-translational modifications are thought to have a critical impact on interaction topologies of histone proteins that serve to regulate transcription and the results presented here suggest cross-linking and mass spectrometry may hold a prominent role in future studies in these and many other areas. Therefore these results represent an important advancement in the application of cross-linking technology, provide new information on modification status and topologies of proteins in cells, and suggest many future applications exist for these technologies.

EXPERIMENTAL PROCEDURES

Synthesis of PIR Cross-linker—The PIR cross-linker Biotin Aspartate Proline (BDP) synthesis was accomplished using solid phase peptide synthesis using an Endeavor 90 system (APPTec, Louisville, KY) employing 9-fluorenylmethoxycarbonyl (Fmoc) chemistry. The super acid sensitive SASRIN-glycine resin was used for the solid support. Amino acid additions to the resin occur in sequential order and are the following, Fmoc-Lys (Biotin), Fmoc₂-Lys, Fmoc-Pro, Fmoc-Asp (t-BOC), and succinic anhydride. Reaction yield at each coupling step was monitored via the absorbance of released FMOC at 307 nm with a cumulative measured yield of >90%. The activated n-hydroxyphthalimide (NHP) ester form of the cross-linker is synthesized in a final esterification step immediately prior to use with TFA-NHP. NHP esters are analogous to the more commonly applied n-hydroxysuccinimide (NHS) esters and were first described in regard to chemical cross-linking by Bich *et al.* (23). Cleavage of BDP from the solid support and removal of the *N-tert*-butoxycarbonyl (t-BOC) protecting groups from the Asp side chains was performed simultaneously using 95% trifluoroacetic acid 5% dichloromethane. Purification of BDP was performed via diethyl ether precipitation using 1:15 (cleavage mixture: diethyl ether). Diethyl ether solution was centrifuged at 3400 g to pellet precipitate. Diethyl ether was decanted and pellet was dried to yield BDP-NHP ester. The final product was verified via direct infusion ESI-MS analysis with a spectrum included as [supplemental Fig. S1](#). BDP was dissolved in dimethyl sulfoxide to a concentration of 500 mM and used immediately or stored at -80°C .

Cell Culture and Cross-linking—HeLa cells were grown at 37°C under a humidified atmosphere containing 5% CO_2 in Dulbecco's modified Eagle's medium containing 10% fetal bovine serum and 1% penicillin/streptomycin until they reached 80% confluence. Cells were harvested by trypsinization and collected into centrifuge tubes. The cells were pelleted and washed 5 times with 1 ml phosphate-buffered saline (PBS) before cross-linking. The cell pellet was resuspended in 170 mM Na_2HPO_4 pH 7.4 and BDP-NHP was added to the suspension with a final concentration of 10 mM. The reaction was carried out at room temperature for 1 h followed by washing 5 times with 1 ml 0.1 M tris buffer pH 8.0. The cells were lysed by heating to 95°C in 4% sodium dodecylsulfate (SDS), 0.1 M tris buffer at pH 8.5. The sample viscosity was reduced by sonication using a GE 130 ultrasonic processor. The sample was centrifuged at 16,000 g for 10 min to pellet any remaining insoluble material. It was then added to a 30 kDa molecular weight cut-off filter (Millipore, Billerica, MA) and concentrated by centrifugation at 7,500 g for 30 min. The sample was washed extensively with 8.0 M urea in 0.1 M tris buffered to pH 8.0 until SDS was completely removed. The protein solution was extracted from the filter and the protein concentration was measured using a Coomassie Plus assay (Pierce, Rockford, IL). The sample was adjusted to a protein concentration of 2 mg/ml and a urea concentration of less than 1 M by adding 100 mM ammonium bicarbonate. Disulfide bonds were reduced with 5 mM tris(2-carboxyethyl) phosphine hydrochloride for 30 min at room temperature, followed by alkylation with 10 mM

iodoacetamide. Proteins were then digested at 37 °C for 16 h with a 1:200 ratio of sequencing grade modified trypsin (Promega, Madison, WI). The peptide sample was de-salted using C18 Sep-Pak solid phase extraction (Waters Corporation, United Kingdom) prior to strong cation exchange (SCX) fractionation of the sample using Poly-SULFOETHYL aspartamide SCX spin columns (Nest Group Inc., Southborough, MA) and a buffer system consisting of ammonium acetate in, 25% acetonitrile containing 0.5% formic acid for elution. Fractions were collected at 0, 50, 80, 300, 500, and 1000 mM ammonium acetate. Prior to affinity enrichment each fraction was desalted using C18 Sep-Pak 50cc (Waters Corporation, United Kingdom). The acetonitrile was removed from the SCX fractions by vacuum centrifugation using an EZ2-Plus evaporator (Genevac, Gardiner, NY) and they were pH adjusted to 7.5 with 1 M NaOH. Biotin containing, PIR labeled peptides were enriched using monomeric avidin immobilized on UltraLink resin (Pierce, Rockford, IL). To each fraction 300 μ l of settled avidin resin was added in 500 μ l of 100 mM ammonium bicarbonate. Samples were washed extensively with 100 mM ammonium bicarbonate to remove unlabeled peptides. Peptides bound to the avidin resin were eluted using 70% acetonitrile containing 0.5% trifluoroacetic acid. Enriched cross-linked peptide samples were concentrated using vacuum centrifugation, and stored at -80 °C until LC-MS analysis.

In addition to avidin enrichment on the PIR labeled peptides as described above, enrichment on the protein level was performed using 1 mg of total protein from PIR labeled HeLa cells. PIR labeled proteins were captured with immobilized monomeric avidin and unlabeled proteins washed away with 100 mM ammonium bicarbonate. PIR labeled proteins were eluted using 8 M urea containing 2 mM D-biotin. The enriched PIR labeled protein sample was then diluted 1:10 with 100 mM ammonium bicarbonate before trypsin digestion as described above. The peptide sample was de-salted using C18 Sep-Pak solid phase extraction, concentrated using vacuum centrifugation, and stored at -80 °C until LC-MS analysis.

Liquid Chromatography-Mass Spectrometry Analysis—Enriched PIR peptide samples were analyzed using a nanoAcquity UPLC (Waters, Milford, MA) coupled to a Velos-FTICR Ultra hybrid mass spectrometer (Thermo Scientific, Waltham, MA). Peptide samples were loaded onto a trap column (3 cm \times 100 μ m i.d.) packed with 200 Å Magic- C4AQ (Michrom Bioresources, Auburn, CA) using a flow rate of 2 μ l/min of 99% solvent A (H₂O containing 0.1% formic acid) and 1% solvent B (acetonitrile containing 0.1% formic acid) where they were washed for a total of 10 min. Peptides were then eluted from the trap column and separated by reversed-phase chromatography over a pulled tip, fused silica analytical column (60 cm \times 75 μ m i.d.) packed with 100 Å Magic- C4AQ at a flow rate of 300 nL/min using a linear gradient from 90% solvent A/10% solvent B to 60% solvent A/40% solvent B over 120 min for a 2 h data acquisition or 240 min for a 4 h data acquisition. Peptides eluting from the column were ionized using electrospray ionization with a spray voltage of 1.8 kV. Peptide ions were analyzed with the Velos-FTICR mass spectrometer, which was operated using a customized method termed Real-time Analysis for Cross-linking Technology (ReACT) (19). The structure of a ReACT method consists of the following mass spectrometry data acquisition parameters. First a high-resolution acquisition (50,000 resolving power (RP) @ 400 m/z) of precursor m/z is performed in the ICR cell. This is followed by a high resolution MS² acquisition on the CID generated fragment ions from the most intense precursor ion with a charge state $\geq 4+$. Settings for MS² include an isolation width of 2 m/z , normalized collision energy of 35V, activation time of 10 ms and 25,000 RP @ 400 m/z . The method then performs a check if any of the masses observed in the MS² spectrum satisfy the following PIR mass relationship equation within a user definable mass tolerance, which was set to 20 ppm:

$$\text{Precursor} = \text{Reporter} + \text{Peptide}_1 + \text{Peptide}_2 \quad (\text{Eq. 1})$$

If the equation was satisfied, a series of four low resolution ion trap MS³ acquisitions were performed to acquire fragmentation spectra of peptides observed in the PIR cross-linked relationships. Settings for MS³ events include acquisition on both the 1+ and 2+ charge states of the peptides found in PIR relationships, an isolation width of 2 m/z , normalized collision energy of 35V, and an activation time of 10 ms. Dynamic exclusion is used with the following parameters: repeat count of 2, repeat duration of 15 s, dynamic exclusion list size of 500, dynamic exclusion duration of 30 s. The monoisotopic precursor selection feature was enabled, however FT preview mode and predictive automated gain control were not.

Peptide samples from PIR enriched proteins were analyzed using the same LC-MS system described above. Peptide samples were loaded onto a trap column (3 cm \times 100 μ m i.d.) packed with 200 Å Magic- C18AQ (Michrom Bioresources, Auburn, CA) using a flow rate of 2 μ l/min of 99% solvent A (H₂O containing 0.1% formic acid) and 1% solvent B (acetonitrile containing 0.1% formic acid) where they were washed for a total of 10 min. Peptides were then eluted from the trap column and separated by reversed-phase chromatography over a pulled tip, fused silica analytical column (60 cm \times 75 μ m i.d.) packed with 100 Å Magic- C18AQ at a flow rate of 300 nL/min using a linear gradient from 95% solvent A/5% solvent B to 60% solvent A/40% solvent B over 120 min. Data dependent analysis (DDA) with the Velos-FTICR mass spectrometer consisted of a high resolution (50,000 RP) MS¹ scan followed by low resolution MS² analysis on the 20 most intense precursors. MS² settings included a normalized collision energy of 35V, isolation width of 2, activation time of 10 ms, activation Q of 0.25 and a minimum signal threshold of 10,000. Charge state exclusion was applied for singly charged precursor ions and those with undetermined charge state. Dynamic exclusion was enabled with settings including an exclusion window of 0.5 m/z low to 1.5 m/z high, exclusion time of 45 s, a list size of 500, and a repeat count of 1.

Data Analysis—The raw mass spectrometry data was converted to mzXML format using ReAdW (Ver. 4.3.1). Data generated from the DDA analysis of the tryptic digest from enriched PIR labeled proteins was used to construct a stage 1 protein database (24) consisting of 3348 proteins. SEQUEST (version UWPR2012.01.3) was used to search the data against the UniProt/SwissProt Human protein database containing both forward and reverse protein sequences (<http://www.uniprot.org> database download date 05/11/12, containing 40,486 total protein sequences). SEQUEST search parameters included a 25 ppm precursor mass tolerance and a 0.36 Da fragment tolerance (0.11 Da fragment offset), trypsin was used as the digesting enzyme, up to 3 missed cleavages, with carbamidomethylation (57.021464 Da) on C as a fixed modification, and oxidation (15.9949 Da) on M as a variable modification. Resulting peptide matches were filtered to a 1% false discovery rate (FDR), which was determined as the ratio of reverse peptide hits to forward peptide hits, resulting in a total of 15,415 unique peptides from 3348 proteins.

For data from ReACT analysis of enriched PIR cross-linked peptide samples, SEQUEST was used to search the data against the stage 1 protein database containing both forward and reverse sequences from those proteins identified as potentially PIR labeled (6696 total protein sequences). Search parameters included 25 ppm precursor mass tolerance and a 0.36 Da fragment tolerance (0.11 Da fragment offset), trypsin was used as the digesting enzyme, up to 3 missed cleavages, fixed modifications included: carbamidomethylation (57.021464 Da) on C, variable modifications included: BDP-stump mass (197.032422 Da) on K and protein N terminus, methylation (14.01560 Da) on K and R, dimethylation (28.033 Da) on K and R, trimethylation (42.046950 Da) on K, and acetylation (42.01565 Da) on

K. Resulting peptide matches were filtered at a static 5% FDR. Peptides, which passed the 5% FDR cutoff were accepted and mapped back to the identified PIR cross-linked relationship. Peptides mapped back to cross-linked relationships were also required to contain a BDP-stump modification at the protein N terminus or a BDP-stump modification at an internal missed cleavage site at K. The presence of additional variable modifications including mono-, di-, tri-methylation, and acetylation were verified against known PTM locations annotated in UniProt. Protein assignment from the peptides that were mapped to cross-linked relationships was performed to maximize intraprotein cross-links for cases where peptides were common to more than one protein sequence. All amino acid residue numbers referred to in the text, tables and figures correspond to the protein sequences in the UniProt databases starting with the initial $M = 0$. Disorder prediction for heat shock protein 90-beta was performed using VSL2 disorder prediction algorithm (25). Homology based molecular modeling of prohibitin and prohibitin-2 was performed using Phyre2 (26) in normal modeling mode. Docking of the resulting monomer models was performed using PatchDock (27) with the prohibitin submitted as the receptor and prohibitin-2 submitted as the ligand with RMSD set to 4 and complex type set to default. Distance constraints from cross-linking results were applied so that the alpha carbons from the cross-linked residues (prohibitin K201 & prohibitin 2 K215) had to be within a distance of less than 35 Å.

Confocal Microscopy—For confocal microscopy samples, HeLa cells were cultured as described above in 35-mm Petri dishes with number 1.5 coverglass bottom (Mat Tek, Ashland, MA). When the cells reached 80% confluence they were washed five times with PBS buffer and reacted with 1 mM PIR cross-linker for 1 h. at room temperature. After the cross-linking reaction, cells were again washed 5 times with 2 ml PBS and fixed by addition of 10% formalin for 10 min at room temperature. Following fixation, cells were incubated with 0.1% triton X-100 in 1 ml PBS for 10 min. The cells were then incubated with 1 µg/ml NeutrAvidin OR green 488 (Invitrogen, Grand Island, NY) in 1 ml PBS containing 0.1% triton X-100 for 1 h. in the dark with constant shaking. Cells were then washed three times with 2 ml PBS followed by incubation with 1 µg/ml propidium iodide for 10 min in PBS. Confocal fluorescent imaging was performed in the red and green fluorescent channels using a Nikon A1R confocal microscope using a 60 × water immersion objective. Images are included as [supplemental Fig. S2](#).

RESULTS

Application of PIR Technology to HeLa Cells—Cross-linker molecules used in PIR technology have several engineered features, which aid in the successful identification of cross-linked sites: a biotin affinity tag to allow for enrichment of low abundance cross-linked species, two low energy CID cleavable bonds to release cross-linked peptides and allow for independent sequencing, and a reporter ion to indicate the presence of a cross-linked product. All experiments presented have applied the BDP-NHP PIR molecule to cross-link HeLa cells. The chemical structure of BDP-NHP is shown in [supplemental Fig. S1](#). After cross-linking we employed sample preparation involving multiple modes of separation including ultrafiltration, strong cation exchange chromatography (SCX), and affinity chromatography. Additionally we used a recently developed dynamic mass spectrometric method (19), which focuses the analytical power of the mass spectrometer specifically on ions of cross-linked peptide pairs (*i.e.*, those that are charge state 4+ or greater and satisfy the PIR mass

relationship in Equation 1). The complete experimental workflow is illustrated in Fig. 1.

For *in vivo* cross-linking and study of proteins other than membrane surface proteins, cell penetration of the cross-linker is important. The biotin group on PIR molecules provides a useful handle to perform assays and determine molecular penetration into cells. Using gold-coupled nanoparticle antibodies and electron microscopy, previous Rink-based PIR molecules were shown to penetrate and react with proteins in the cytosol of Gram-negative bacteria (17, 28). To obtain complimentary verification of the membrane permeability of PIR molecules used with HeLa cells in the present study, we used fluorescent confocal microscopy. Confocal images of fluorophore-coupled avidin on PIR-reacted HeLa cells illustrated PIR penetration into cytoplasm and nuclear regions and labeled sites on intracellular proteins including nuclear proteins is included in [supplemental Fig. S2](#).

Here we applied a two stage approach used previously in the application of PIR to *Shewanella oneidensis* (24, 29). The first stage consists of enrichment and shotgun proteomics identification of PIR labeled proteins. In this stage, we identified 15,415 unique peptides at less than 1% FDR, corresponding to 3348 proteins that are putative reactive targets with the PIR cross-linker ([supplemental Table S1](#)). The second stage consists of affinity enrichment of PIR labeled peptides allowing for the identification of the cross-linked site of interaction. A unique feature of PIR technology is that identification of each peptide in a cross-linked complex proceeds independently. Peptide mass determination and fragmentation spectral acquisition events and subsequent database searches allow each peptide to be identified independent of the other linked peptide. Furthermore, each identification event is also evaluated against a reverse sequence database so that every peptide sequence can be selected above a chosen FDR threshold. Application of these techniques to human cells resulted in 368 identified cross-linked peptide pairs at $\leq 5\%$ FDR. The 5% FDR threshold refers to setting an E-value threshold on the peptide assignments from a SEQUEST search of the MS³ spectra against the UniProt human database containing forward and reverse protein sequences such that 5% of the identified peptides passing the E-value threshold result from a match to a reverse sequence. A table of these 368 cross-linked peptide pairs including observed peptide masses, peptide sequences, and protein descriptions is included in [supplemental Table S2](#). Annotated fragment ion spectra for each of the peptides in these 368 cross-linked peptide pairs are included as [supplemental Data Set S1](#). In addition to the 368 cross-linked peptide pairs for which both peptides were identified at $\leq 5\%$ FDR, the data set presented here also included 532 additional cross-linked peptide pairs for which only one peptide was identified at less than 5% FDR but the second peptide was identified greater than 5% FDR. The peptides with less confident identifications ($>5\%$ FDR) were assigned to the top scoring peptide se-

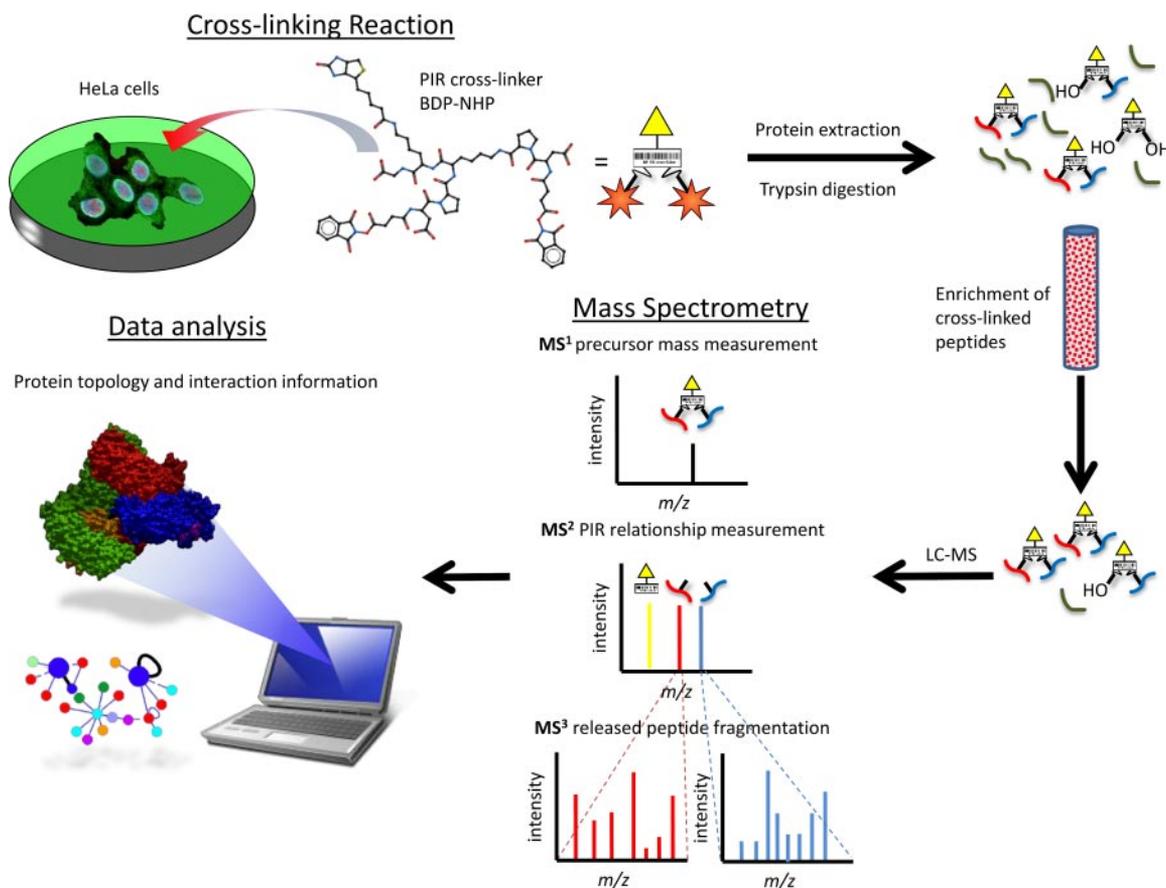


FIG. 1. Experimental overview: Sample preparation, PIR cross-linking and mass spectrometry analysis, and data analysis.

quence identified from a SEQUEST search matching both in accurate precursor mass and greatest number of fragment ions. It is important to note that although high quality fragmentation information was not obtained for one of the released peptides from these cross-linked peptide pairs their masses were still measured with high mass accuracy and contain a BDP modified internal lysine residue. These data are included in [supplemental Table S3](#) and highlight a persistent challenge for all cross-linking studies in that high quality fragment spectra are required for both peptides to yield confident cross-linked peptide pair identification. This is one area in particular where future improvements to mass spectrometry methods and informatics will help overcome the challenges faced with cross-linking experiments. Additionally improvements to cross-linker chemical design that would produce released peptides of primarily charge state 2+ and 3+, along with application of different digestion enzymes could contribute to overcoming challenges in this area. By combining these 532 cross-linked peptide pairs with higher confidence set of 368 cross-linked peptide pairs and filtering for redundancy yields a total of 783 unique cross-linked peptide pairs. The mean mass error for the PIR relationships for these 783 cross-linked peptide pairs was 2.9 ppm with over 84% (664) measured at less than 5 ppm mass error as can be seen in the histogram included in [supplemental Fig. S3](#).

The data were further analyzed using a recently developed online software tool and database for cross-linking results named XLink-DB (20). XLink-DB automates several important analyses for large scale cross-linking data sets including generating an interaction network view, comparing observed interactions to known protein interaction databases, and mapping of cross-links onto known structural data. Fig. 2A illustrates a protein interaction network generated from the 648 unique cross-linked peptide pairs from human cells. The network consists of 307 nodes representing the identified cross-linked proteins connected by 446 edges representing observed intraprotein and interprotein cross-links. Highly connected hub nodes are highlighted with a larger node size and their UniProt identifier. Major hubs include histone proteins, ribosomal proteins, and heterogeneous ribonucleoproteins. An interactive version of the high confidence cross-links within this network is available online at (<http://brucelab.gs.washington.edu/CrossLinkDB/>). Importantly, such a protein interaction network generated using chemical cross-linking contains a depth of information beyond what similar protein interaction networks generated by affinity pull-down methods contain. In addition to the identities of the interacting proteins, topological information about the interacting regions of these proteins is contained in the cross-linked residues. It is worth noting that for cross-linked peptides to be observed by the

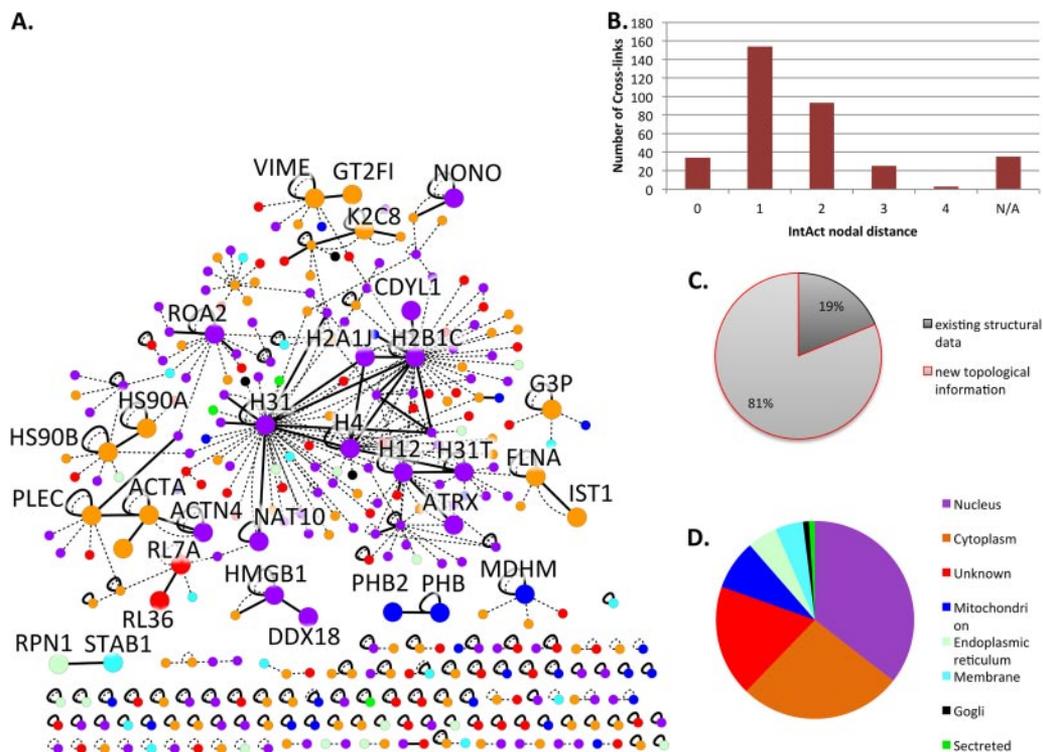


FIG. 2. A, Protein interaction network generated exclusively from cross-linking results. Network consists of 307 nodes representing proteins connected by 446 edges representing observed intraprotein and interprotein cross-links. Nodes are colored according to subcellular localization with major hubs indicated by larger node size. Bold black edges indicate cross-links for which both peptides were identified at less than 5% FDR whereas thin dashed edges are cross-links for which only one peptide passed the FDR threshold. **B,** Distribution of nodal distance generated using XLink-DB (<http://brucelab.gs.washington.edu/CrossLinkDB/>) to compare protein interactions from network in A with protein-protein interactions in the IntAct database. **C,** Pie chart indicating cross-links that can be mapped to existing structures in the PDB and those providing new topological information. **D,** Pie chart indicating subcellular localization of cross-linked proteins.

approach described in this paper must have existed in relatively close proximity and orientation to one another billions to trillions of times (assuming femtomole to picomole sensitivity levels for peptides using modern mass spectrometers) during the reactive life-time of the cross-linker ($\lambda \sim 8$ min in neutral pH aqueous buffer). Thus what could be viewed as limitations of current cross-linking technology actually provide a valuable benefit in that nonspecific protein-protein interactions, which are a commonly an Achilles' heel for AP-MS approaches (30), are likely to be less frequently detected because the linkage takes place on proteins within their native environment. Additionally, linkage of only two or a few specific lysine residues in two protein sequences indicates that the two proteins are close to one another in cells with a specific relative orientation so as to allow only specific cross-links to be formed. These two features: 1) high frequency of being close to one another and 2) high frequency of being close to one another with a specific orientation, are hallmarks of specific protein interactions. Therefore, chemical cross-linking technologies can provide this level of detail and will eventually be exploited more effectively to help better understand specific protein interactions in cells.

The protein interaction network was processed with XLink-DB to compare the protein interactions discovered by

cross-linking with previously known interactions present in the IntAct database (31). A histogram of nodal distances between discovered and known interactions is displayed in Fig. 2B. The majority interprotein cross-links have a distance of either one or two, meaning that the two cross-linked proteins are not known to directly interact in the IntAct database but they either share a common interacting partner, which connects them, or their interacting partners are known to directly interact. Although there are only 34 cross-links that have an IntAct nodal distance of zero, it should be noted that there are several examples of well-known interacting proteins that do not exist as direct interacting partners in the IntAct database. For example the interaction between histones H31 (UniProt accession P68431) and H4 (UniProt accession P62805) are well established interacting partners in the nucleosome complex however have a nodal distance of one according to IntAct. Such cases are inevitable reflections of incomplete knowledge and annotations in protein interaction databases. Despite these instances, protein interactions with nodal distances of 1 or greater potentially represent newly discovered protein interactions. For example, the E3 ubiquitin-protein ligase CBL-B where lysine 468 was cross-linked with lysine 29 of the high mobility group protein B1. These two proteins have a nodal distance of one in IntAct, however are

observed as direct interacting partners in this study. As described in the discussion, cross-links identify the endoplasmic reticulum membrane protein dolichyl-diphosphooligosaccharide-protein glycosyltransferase subunit 1 (aka ribophorin-1) and the outer membrane glycoprotein stabilin-1 as direct interacting partners although they are separated by two nodes in IntAct.

Nodes in the network are colored according to their top subcellular location obtained from the UniProt database. As expected for many of the nodes, interactions are discovered between proteins from the same subcellular compartment. However, there are also interactions between proteins from different subcellular compartments, which are readily explainable. For example interactions between nuclear and cytoplasmic proteins are to be expected, with many reports of proteins moving between these compartments through the nuclear pore 20 (32). It is also worth noting that the majority of proteins have multiple entries in UniProt for subcellular location so although the top entries for two cross-linked proteins may not match, subsequent entries may overlap. For example, the cross-link between alpha actinin 4, which has a top subcellular annotation as nuclear and a secondary annotation of cytoplasmic, and beta actin, which has a top subcellular annotation of cytoplasmic. In addition it is unreasonable to expect that the UniProt annotation for the subcellular localization of proteins is both complete and comprehensive, therefore it is quite possible that two proteins of seemingly different subcellular locations are identified as cross-linked. Fig. 2C illustrates the percentage of proteins from the various subcellular locations in a pie diagram. Importantly all major subcellular locations are represented and the large percentage of cytosolic and nuclear proteins indicates adequate penetration of the PIR cross-linker into cells, a conclusion supported by the fluorescence microscopy data in [supplemental Fig. S2](#).

By attempting to map the 691 unique cross-linked sites from 783 cross-linked peptides to available x-ray crystal structures in the Protein Data Bank, Euclidean distances between the linked alpha carbon atoms were obtained for 130 cross-links. The crystal structures with cross-links mapped are available for viewing online at (<http://brucelab.gs.washington.edu/CrossLinkDB/>). The measured distances spanned a range from 5.1 to 54.3 Å with a median distance of 14.9 Å as can be seen in a histogram in [supplemental Fig. S4](#). The distances on average are about 20 Å less than the maximum spacer arm length for BDP-PIR (~35 Å) with ~95% of the total measured distances less than 35 Å. The seven cases where measured distances exceeded 35 Å can be rationalized by considering factors such as the flexibility of protein structures in solution. For example the largest distance mapped (54.3 Å) corresponds to a cross-link between K37 of H3 and K91 of H4. Being that H3K37 is located on the very flexible N-terminal tail of H3 it is possible the actual distance between the cross-linked sites is shorter than that measured from the

crystal structure. It is also possible that this cross-link results from the linkage between H3 and H4 from stacked nucleosome particles, rather than within a single nucleosome complex. As an example of another explainable case; the cross-link between K488 and K797 of DNA topoisomerase 2-alpha was mapped as an intraprotein cross-link with a distance of 42.2 Å, however it is possible this cross-link may span a shorter distance existing between two identical subunits of DNA topoisomerase 2-alpha being that it is known to form a homodimer (33). The PDB structure (1LWZ) used to map the distance of the cross-linked site for this protein only contains a monomeric structure so we were only able to map this distance as an intralink. The observed distances match well with other studies in our laboratory applying PIR cross-linking in *E. coli* (17–19). These distances also appear consistent with those observed and/or predicted in other studies that employed cross-linkers with much smaller linker arms such as DSS or BS3 (spacer arm length ~11.4 Å) (5, 11, 16, 34). For example Herzog *et al.* measured the median Euclidean distances between alpha carbons for 70 interprotein and 287 intraprotein DSS cross-linked peptide pairs from protein phosphatase 2A complexes to be 19.6 Å and 15.4 Å respectively (11). These measurements suggest that factors other than cross-linker length play a role in the determination of which sites are observed from large-scale cross-linking studies. The cross-links, which we were not able to map onto crystal structures, provide valuable new information on interaction topologies for many proteins that have no existing and/or partially resolved crystal structures. As can be seen by the pie chart in Fig. 2C no structural information exists in the Protein Data Bank for most (over 80%) of the total cross-linked sites from both the high and low confidence data sets identified in this study. For example cross-links observed in the highly disordered N-terminal tails of the core histone proteins, which are absent from nucleosome crystal structures. Additionally cross-links membrane proteins prohibitin 1 (PHB) and prohibitin-2 (PHB2) can provide new insights into the topology of interaction for these proteins. These and other examples will be discussed in more detail in the discussions section.

Of the 368 high confidence cross-linked pairs, 284 consisted of two peptides from within the same protein sequence, meaning they are either intraprotein linkages or interprotein linkages from homo-multimers. These two types are not easily distinguished except for cases where the two peptides are exactly the same sequence (peptide homodimer) or share some overlapping sequence, which only occurs once per protein molecule. There are 12 such unambiguous homodimers in the present data set.

The high number of observed intraprotein and homodimer cross-links is to be expected for several reasons. First, intraprotein cross-links are formed in greater abundance because of the fact that once one reactive group of the cross-linker reacts with a protein molecule the second functional group

becomes tethered and is constrained to react with a free amine site nearby, often times within the same molecule. Furthermore, self-interacting proteins are anticipated to be a predominant type of specific protein-protein interaction because of colocalization and a relatively high local concentration of binding partners (35, 36). Select examples of unambiguous homodimer cross-links are discussed in more detail below in the discussion section.

In addition to enabling identification of traditional cross-linked peptides, the data presented here also demonstrate the ability to identify cross-linked peptides containing additional post-translational modifications including methylation, and dimethylation on lysine and arginine, trimethylation lysine, and acetylation on lysine. It has been previously noted that cross-linked sites observed in *E. coli* were also sites of lysine acetylation (21). This raises interesting questions about the relative reactivity of these particular lysine residues as well as the influence of these and nearby lysine sites in defining protein topology and regulation of protein interactions. It seems plausible that lysine residues, which are targets of post-translational modification, reside on the surface of the protein to increase accessibility. These specific residues also appear to represent local “hot spots” of reactivity for modifying enzymes as well as cross-linker molecules. The application of chemical cross-linking to understand the impact of post-translational modifications on protein topology and interactions is currently uncharted territory, but could greatly accelerate understanding of the relevance of post-translational modifications in biological systems. A primary factor that has inhibited this advance is the large increase in database search space when allowing for the possibility of post-translational modifications that is further exacerbated by the N^2 increase in search space encountered when attempting to assign two peptide sequences from a single precursor mass (37). Therefore confident identification of variable post-translational modifications from complex samples becomes impractical, if not intractable, when working with traditional, non-cleavable cross-linkers. The cleavable feature of PIR cross-linkers allows for individual accurate mass measurements to be made on the released peptides, eliminating the N^2 increase in search space, and allowing for the confident identification of variable post-translational modifications. The possibility of identifying post-translational modifications in the cross-linked peptide data set from HeLa cells was investigated. Excitingly, confident identification was achieved on 93 unique cross-linked peptide pairs, which contained additional post-translational modifications including mono-, di-, and tri-methylation on Lys as well as acetylation on Lys residues (supplemental Table S2). These 93 cross-linked peptides contain 21 unique sites of modification on 13 different proteins. Importantly these data are the first reported cross-linked peptides containing *in vivo* post-translational modifications known to be important for regulating protein topology and interactions and having a direct impact on protein function. For example we identified

82 cross-linked peptide pairs from histones, which also contained additional post-translational modifications. All of the observed histone cross-linked sites and modifications observed are included in Table I. These data, discussed in further detail below, provide unique insight into the structure of histone proteins and how their topology changes with various modification states. It is important to note that the lysine side chains linked by our cross-linker must be unmodified because the activated ester reactive groups will not react with acetylated or methylated amines. Furthermore, it is worth noting that six peptides were assigned to have modified Lys or Arg residues as their C-terminal residue. Although there are reports of trypsin cleaving at methylated Lys, there is a possibility these represent incorrect assignments because of the lack of specificity of trypsin to cleave at modified Lys or Arg. However six peptides out of 736 total peptides in our high confidence set of cross-links corresponds to $\sim 0.8\%$ of peptide identifications, well below our 5% FDR threshold.

DISCUSSION

Unambiguous Cross-linked Homodimers—If one accepts the theory that protein colocalization lies at the origin of all protein-protein interactions and that most interactions between paralogs evolved from ancestral homodimer interactions, then understanding topologies of interaction between homodimers is at the heart of understanding how and why protein molecules interact with one another (35, 36). Because of their importance, homo-oligomeric interactions are of intense interest for drug development effort for the treatment of a myriad of human diseases including cancer and HIV. HSP90 is one such homodimer that has significant clinical significance in cancer (38). One example of an unambiguous homodimer cross-link is the peptide **FYEAFSK⁴³⁴NLK** spanning residues 428–437 (bold underline indicates cross-linked residue) from heat shock protein 90 -beta (HS90B). The mass spectra identifying the HS90B homodimer cross-link are shown in Fig. 3A–3C. HSP90 proteins are highly conserved, essential molecular chaperones that assist in the proper folding and stabilization of proteins as well as regulation of cellular signaling pathways. The location of the homodimer cross-link was mapped onto the homologous structure for HS90 homodimer from yeast (39). The identified cross-linked site lies near the transition between the catalytic protein binding domain of HS90 that serves to bind substrates and contains the catalytic loop, and the C-terminal dimerization domain Fig. 3D. Prediction of protein disorder with the VSL2 disorder predictor (25) using the sequence from HS90B indicates that K434 is located near a transition from ordered to disordered structure, which appear to be more susceptible to cross-linker reaction in large scale studies (21). Five additional cross-links were observed in HS90B including two with K434 (K434-K606, and K346-K434) linking this site with sites in the C-terminal dimeriza-

TABLE 1

Histone cross-links with modifications. *Identified cross-linked peptide sequence with mass of modifications indicated in brackets following the modified residue. Modifications are: 142.11-mono-methylation, 156.13-di-methylation, 170.11-acetylation, 170.14-tri-methylation, 325.13-BDP stump mass indicating cross-linked residue. † Modification checked against data contained in UniProt for known modification sites. ^ amino acid residue number for cross-linked site starting with initial Met = 1.

peptide 1*	peptide 2*	Protein 1 (UniProt)	Protein 2 (UniProt)	peptide 1 modification verification†	peptide 2 modification verification†	pep1 xlink site	pep2 xlink site
K[142.11]STGGK[325.13]APR	TK[325.13]QTAR	sp P68431 H31_HUMAN +3	sp P68431 H31_HUMAN +3	H3K9me, Uniprot		14	4
K[156.13]STGGK[325.13]APR	TK[325.13]QTAR	sp P68431 H31_HUMAN +3	sp P68431 H31_HUMAN +3	H3K9me2, Uniprot		14	4
K[142.11]STGGK[325.13]APR	K[325.13]QLATK	sp P68431 H31_HUMAN +3	sp P68431 H31_HUMAN +3	H3K9me, Uniprot		14	18
K[156.13]STGGK[325.13]APR	K[325.13]QLATK	sp P68431 H31_HUMAN +3	sp P68431 H31_HUMAN +3	H3K9me2, Uniprot		14	18
K[156.13]STGGK[325.13]APR	K[325.13]QLATK[170.11]JAAR	sp P68431 H31_HUMAN +3	sp P68431 H31_HUMAN +3	H3K9me2, Uniprot	H3K23ac, Uniprot	14	18
K[156.13]SAPATGGVK[325.13]KPHR	YQK[325.13]STELLR	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +2	H3K27me2, Uniprot		36	56
K[156.13]SAPATGGVK[156.13]K[325.13]PHR	YQK[325.13]STELLR	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +3	H3K27me2, H3K36me, Uniprot		37	56
K[156.13]SAPATGGVK[325.13]KPHR	K[325.13]QLATK	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +3	H3K27me2, Uniprot		36	18
K[142.11]STGGK[325.13]APR	K[325.13]QLATK[170.11]JAAR	sp P68431 H31_HUMAN +3	sp P68431 H31_HUMAN +2	H3K9me, Uniprot	H3K23ac, Uniprot	14	18
K[156.13]SAPATGGVK[142.11]K[325.13]PHR	YQK[325.13]STELLR	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +3	H3K27me2, H3K36me, Uniprot		37	56
K[156.13]SAPATGGVK[325.13]K[142.11]PHR	YQK[325.13]STELLR	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +3	H3K27me2, H3K36me, Uniprot		36	56
K[156.13]SAPATGGVK[325.13]KPHR	K[325.13]QLATK[170.11]JAAR	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +2	H3K27me2, Uniprot	H3K23ac, Uniprot	36	18
K[156.13]STGGK[325.13]APR	YQK[325.13]STELLR	sp P68431 H31_HUMAN +3	sp P68431 H31_HUMAN +3	H3K9me2, Uniprot		14	56
K[156.13]SAPATGGVK[325.13]KPHR	TK[325.13]QTAR	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +3	H3K27me2, Uniprot		36	4
K[156.13]STGGK[325.13]APR	RGVVK[325.13]R	sp P68431 H31_HUMAN +3	sp P62805 H4_HUMAN	H3K9me2, Uniprot		14	44
K[156.13]SAPATGGVK[156.13]K[325.13]PHR	TK[325.13]QTAR	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +3	H3K27me2, H3K36me2, Uniprot		37	4
K[156.13]SAPATGGVK[325.13]K[156.13]PHR	K[325.13]QLATK	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +3	H3K27me2, H3K36me2, Uniprot		37	4
K[156.13]SAPATGGVK[325.13]K[156.13]PHR	K[325.13]QLATK	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +3	H3K27me2, H3K36me2, Uniprot		36	18
K[156.13]SAPATGGVK[142.11]K[325.13]PHR	K[325.13]QLATK[170.11]JAAR	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +2	H3K27me2, H3K36me, Uniprot	H3K23ac, Uniprot	37	18
K[156.13]SAPATGGVK[156.13]K[325.13]PHR	K[325.13]QLATK[170.11]JAAR	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +2	H3K27me2, H3K36me2, Uniprot		37	18
K[156.13]SAPATGGVK[325.13]K[156.13]PHR	YQK[325.13]STELLR	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +3	H3K27me2, H3K37me2, Uniprot		36	56
K[156.13]SAPATGGVK[325.13]KPHR	K[325.13]TESHHK	sp P68431 H31_HUMAN +2	sp P04908 H2A1B_HUMAN +1	H3K27me2, Uniprot		36	119
K[156.13]SAPATGGVK[156.13]K[325.13]PHR	K[325.13]TESHHK	sp P68431 H31_HUMAN +2	sp P04908 H2A1B_HUMAN +1	H3K27me2, H3K36me2, Uniprot		37	119
K[156.13]SAPATGGVK[142.11]K[325.13]PHR	K[325.13]TESHHK	sp P68431 H31_HUMAN +2	sp P04908 H2A1B_HUMAN +1	H3K27me2, H3K36me, Uniprot		37	119
K[156.13]SAPATGGVK[325.13]KPHR	K[156.13]STGGK[325.13]APR	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +3	H3K27me2, Uniprot	H3K9me2, Uniprot	36	14
K[142.11]STGGK[325.13]APR	K[170.14]SAPATGGVK[325.13]KPHR	sp P68431 H31_HUMAN +3	sp P68431 H31_HUMAN +2	H3K9me, Uniprot	H3K27me3, Uniprot	14	36
K[142.11]STGGK[325.13]APR	K[156.13]SAPATGGVK[325.13]KPHR	sp P68431 H31_HUMAN +3	sp P68431 H31_HUMAN +2	H3K9me, Uniprot	H3K27me2, Uniprot	14	36
K[156.13]SAPATGGVK[142.11]K[325.13]PHR	K[156.13]STGGK[325.13]APR	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +3	H3K27me2, H3K36me, Uniprot		37	14
K[156.13]SAPATGGVK[156.13]K[325.13]PHR	K[156.13]STGGK[325.13]APR	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +3	H3K27me2, H3K36me2, Uniprot		37	14
K[156.13]SAPATGGVK[156.13]K[325.13]PHR	K[170.14]STGGK[325.13]APR	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +3	H3K27me2, H3K36me2, Uniprot		37	14
K[156.13]SAPATGGVK[156.13]K[325.13]PHR	RGVVK[325.13]R	sp P68431 H31_HUMAN +2	sp P62805 H4_HUMAN	H3K27me2, H3K36me2, Uniprot		37	44
K[156.13]SAPATGGVK[156.13]K[325.13]PHR	LAHYNK[325.13]R	sp P68431 H31_HUMAN +2	sp P33778 H2B1B_HUMAN +2	H3K27me2, H3K36me2, Uniprot		37	85
K[156.13]SAPATGGVK[325.13]K[156.13]PHR	TK[325.13]QTAR	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +3	H3K27me2, H3K37me2, Uniprot		36	4
K[156.13]SAPATGGVK[156.13]K[325.13]PHR	K[325.13]QLATK	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +3	H3K27me2, H3K36me2, Uniprot		37	18
K[156.13]SAPATGGVK[325.13]KPHR	K[325.13]STGGK	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +3	H3K27me2, Uniprot		36	9
K[142.11]SAPATGGVK[325.13]KPHR	K[156.13]STGGK[325.13]APR	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +3	H3K27me, Uniprot	H3K9me2, Uniprot	36	14
K[156.13]SAPATGGVK[325.13]KPHR	RGVVK[325.13]R	sp P68431 H31_HUMAN +2	sp P62805 H4_HUMAN	H3K27me2, Uniprot		36	44

TABLE 1—Continued

peptide 1*	peptide 2*	Protein 1 (UniProt†)	Protein 2 (UniProt†)	peptide 1 modification verification‡	peptide 2 modification verification‡	pep1 xlink site	pep2 xlink site
K[156.13]SAPATGGVK[142.11]K[325.13]P[HR]	RGGVK[325.13]R	sp P68431 H31_HUMAN +2	sp P62805 H4_HUMAN	H3K27me2, H3K36me, Uniprot		37	44
K[156.13]SAPATGGVK[325.13]K[142.11]P[HR]	RGGVK[325.13]R	sp P68431 H31_HUMAN +2	sp P62805 H4_HUMAN	H3K27me2, H3K36me, Uniprot		36	44
K[156.13]STGGK[325.13]A[PR]	K[156.13]STGGK[325.13]A[PR]	sp P68431 H31_HUMAN +3	sp P68431 H31_HUMAN +3	H3K9me2, Uniprot	H3K9me2, Uniprot	14	14
K[156.13]SAPATGGVK[142.11]K[325.13]P[HR]	TK[325.13]Q[TA]R	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +3	H3K27me2, H3K36me, Uniprot		37	4
K[156.13]SAPATGGVK[142.11]K[325.13]P[HR]	K[325.13]Q[LA]TK	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +3	H3K27me2, H3K36me, Uniprot		37	18
K[156.13]SAPSTGGVK[156.13]K[325.13]P[HR]	YOK[325.13]STELLIR	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +3	H3K27me2, H3K36me2, Uniprot		37	56
K[325.13]STGGK[170.11]A[PR]	TK[325.13]Q[TA]R	sp P68431 H31_HUMAN +3	sp P68431 H31_HUMAN +3	H3K14ac, Uniprot		9	4
K[325.13]Q[LA]TK[170.11]A[PR]	TK[325.13]Q[TA]R	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +3	H3K14ac, Uniprot		18	4
K[170.14]SAPATGGVK[325.13]A[PR]	TK[325.13]Q[TA]R	sp P68431 H31_HUMAN +3	sp P68431 H31_HUMAN +3	H3K9me3, Uniprot		14	4
K[170.14]STGGK[325.13]A[PR]	K[325.13]Q[LA]TK	sp P68431 H31_HUMAN +3	sp P68431 H31_HUMAN +3	H3K9me3, Uniprot		14	18
K[170.14]SAPATGGVK[325.13]P[HR]	YOK[325.13]STELLIR	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +3	H3K27me3, Uniprot		37	56
GLGK[325.13]GGAK[170.11]R	LLLPGELAK[325.13]H[V]SEG[TK]	sp P62805 H4_HUMAN	sp P62807 H2B1C_HUMAN +3	H4K16ac, Uniprot		12	108
K[170.14]STGGK[325.13]A[PR]	K[325.13]Q[LA]TK[170.11]A[PR]	sp P68431 H31_HUMAN +3	sp P68431 H31_HUMAN +2	H3K9me3, Uniprot		14	18
K[170.11]SAPATGGVK[325.13]K[PHR]	YOK[325.13]STELLIR	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +3	H3K27ac, Uniprot		36	56
K[170.14]SAPATGGVK[325.13]K[PHR]	YOK[325.13]STELLIR	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +3	H3K27me3, Uniprot		36	56
K[170.14]SAPATGGVK[325.13]K[PHR]	K[325.13]Q[LA]TK	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +3	H3K27me3, Uniprot		36	18
K[325.13]Q[LA]TK[170.11]A[PR]	TK[325.13]Q[TA]R	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +3	H3K14ac, Uniprot		18	9
K[170.14]SAPATGGVK[325.13]K[PHR]	K[325.13]Q[LA]TK[170.11]A[PR]	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +3	H3K27me3, Uniprot		36	4
K[170.14]SAPATGGVK[325.13]K[PHR]	K[325.13]Q[LA]TK[170.11]A[PR]	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +2	H3K23ac, Uniprot		36	18
K[170.14]SAPATGGVK[325.13]K[PHR]	K[325.13]Q[LA]TK[170.11]A[PR]	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +2	H3K36me, Uniprot		18	27
K[325.13]Q[LA]TK[170.11]A[PR]	K[325.13]SAPATGGVK[142.11]K[PHR]	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +2	H3K27me3, Uniprot		36	119
K[170.14]SAPATGGVK[325.13]K[PHR]	K[170.14]STGGK[325.13]K[PHR]	sp P68431 H31_HUMAN +2	sp P04908 H2A1B_HUMAN +1	H3K27me3, Uniprot		36	14
K[325.13]Q[LA]TK[170.11]A[PR]	K[325.13]SAPATGGVK[156.13]K[PHR]	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +3	H3K27me3, Uniprot		37	18
K[170.14]SAPATGGVK[325.13]K[PHR]	K[156.13]SAPATGGVK[325.13]K[PHR]	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +2	H3K36me2, Uniprot		18	27
K[170.14]SAPATGGVK[325.13]K[PHR]	K[142.11]STGGK[325.13]A[PR]	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +2	H3K27me2, Uniprot		23	36
K[170.14]SAPATGGVK[325.13]K[PHR]	K[156.13]SAPATGGVK[325.13]K[PHR]	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +2	H3K9me, Uniprot		23	14
K[170.14]SAPATGGVK[325.13]K[PHR]	K[156.13]STGGK[325.13]A[PR]	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +3	H3K9me2, Uniprot		23	14
K[325.13]GLGK	K[156.13]STGGK[325.13]A[PR]	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +3	H3K9me2, Uniprot		8	14
KTSHHK[325.13]JAK	K[156.13]STGGK[325.13]A[PR]	sp P62805 H4_HUMAN	sp P68431 H31_HUMAN +3	H3K9me2, Uniprot		125	14
GK[325.13]GGK	K[156.13]SAPATGGVK[325.13]K[PHR]	sp P62805 H4_HUMAN +2	sp P68431 H31_HUMAN +2	H3K27me2, Uniprot		5	36
K[325.13]AAR	K[142.11]SAPATGGVK[325.13]K[PHR]	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +2	H3K27me, Uniprot		23	36
K[325.13]AAR	K[325.13]SAPATGGVK[142.11]K[PHR]	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +2	H3K36me, Uniprot		23	27
K[325.13]AAR	K[156.13]SAPATGGVK[156.13]K[325.13]P[HR]	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +2	H3K27me2, H3K36me2, Uniprot		23	37
K[325.13]AAR	K[156.13]SAPATGGVK[142.11]K[325.13]P[HR]	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +2	H3K27me2, H3K36me, Uniprot		23	37
K[325.13]AAR	K[156.13]SAPATGGVK[325.13]K[142.11]P[HR]	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +2	H3K27me2, H3K37me, Uniprot		23	36
KTSHHK[325.13]JAK	K[156.13]SAPATGGVK[325.13]K[PHR]	sp P04908 H2A1B_HUMAN	sp P68431 H31_HUMAN +2	H3K27me2, Uniprot		125	36
K[325.13]AAR	K[156.13]SAPATGGVK[325.13]K[156.13]P[HR]	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +2	H3K27me2, H3K37me2, Uniprot		23	36
GK[325.13]GGK	GLGK[325.13]GGAK[170.11]R	sp P62805 H4_HUMAN +2	sp P62805 H4_HUMAN	H4K16ac, Uniprot		5	12
K[325.13]AAR	K[170.14]STGGK[325.13]A[PR]	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +3	H3K9me3, Uniprot		23	14
K[325.13]AAR	K[325.13]STGGK[170.11]A[PR]	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +3	H3K14ac, Uniprot		23	9
K[325.13]Q[LA]TK	K[325.13]STGGK[170.11]A[PR]	sp P68431 H31_HUMAN +3	sp P68431 H31_HUMAN +3	H3K14ac, Uniprot		18	9
K[325.13]AAR	K[325.13]SAPATGGVK[170.14]K[PHR]	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +2	H3K36me3, Uniprot		23	27
K[325.13]AAR	K[325.13]SAPSTGGVK[170.14]K[PHR]	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +2	H3K36me3, Uniprot		23	27
K[325.13]AAR	K[170.14]SAPATGGVK[325.13]K[PHR]	sp P68431 H31_HUMAN +2	sp P68431 H31_HUMAN +2	H3K27me3, Uniprot		23	36
KTSHHK[325.13]JAK	K[170.14]STGGK[325.13]A[PR]	sp P04908 H2A1B_HUMAN	sp P68431 H31_HUMAN +3	H3K9me3, Uniprot		125	14

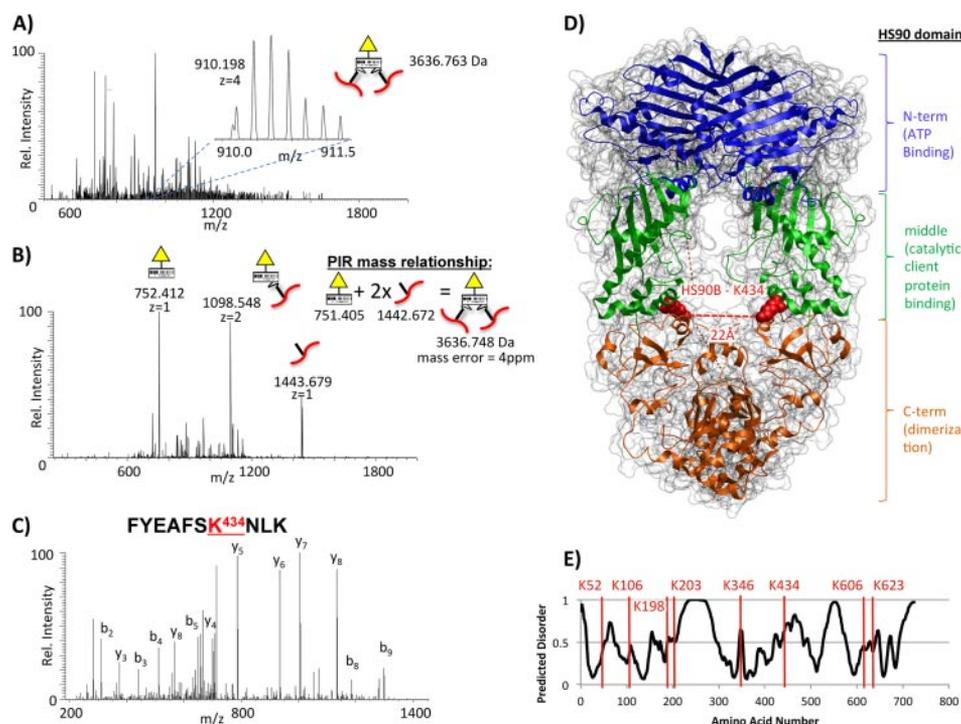


FIG. 3. **A**, Precursor FT-ICR mass spectrum for with inset illustrating the 4+ isotope distribution at m/z 910. 198 for the homodimer cross-linked peptide pair. **B**, High resolution MS² spectrum for cross-linked peptide pair indicating released peptide and reporter ions. **C**, Ion trap MS3 spectrum used to identify peptide FYEAFSKNLK from HS90B. **D**, Crystal structure for the yeast HSP90 dimer (PDB = 2CG9) highlighting position of cross-linked lysine 434 from HS90B near the interface of the middle and C-terminal domains (note: particular lysine residue is conserved between yeast and human although appears as K423 in yeast crystal structure). **E**, Predicted disorder plot (generated using VSL2 disorder prediction algorithm) for HS90B indicating presence of K434 near a transition between order and disordered region.

tion domain and in the disordered amphiphilic loop implicated in client protein interactions respectively (40). Of the three additional HS90B cross-links, two are in the N-terminal region (K52-K106, K198-K203) and one is in the C-terminal region K606-K623. Although these additional cross-links do not provide unambiguous information about the multimeric state of HS90B, they are nonetheless structurally informative. Recently HSP90 proteins have become the target of anticancer treatments because of their stabilization of several oncogenic factors promoting tumor growth (41). Two cytosolic isoforms of HSP90 exist in humans, including HS90B (constitutive expression form) and heat shock protein 90- α (HS90A) (inducible expression form). These two isoforms share 85% sequence homology and are thought to be the result of a gene duplication event that occurred 500 million years ago (42). The two isoforms of HSP90 are thought to exist primarily as homodimers however some evidence for alpha-beta homodimers also exists (43–45). Interestingly, a hetero-dimer cross-linked peptide pair was identified that included the same site of HS90B (K434), and the peptide FYEQFSK⁴⁴²NIK spanning residues 436–445 of HS90A. The respective mass spectra used to identify this hetero-dimer cross-linked peptide pair are illustrated in supplemental Fig. S5. Sequence alignment between HS90A and HS90B from human and HS90 from yeast

is included in supplemental Fig. S6. The sequence alignment illustrates that all of the lysine residues cross-linked in HS90B are conserved (highlighted red) in HS90A and in yeast HS90 except for K204 in yeast and K347 in HS90A, which are both substituted to arginine. As mentioned above the cross-linked residues identified here (HS90B-K434, and HS90A-K442) lie near the interface of the C-terminal dimerization domain and the middle domain of HSP90, which is important for client protein binding and also contains the catalytic loop (40). Interestingly, both K434 and K623 of HS90B, and K442 of HS90A have been identified as acetylation sites (46, 47). Acetylation has been shown to regulate HS90 activity and can inhibit its dynamic association with other chaperones and cochaperones (48). Several studies have correlated HSP90 activity with histone deacetylase (HDAC) activity, suggesting that combination cancer therapy with HSP90 inhibitors and HDAC inhibitors may have a synergistic effect (41). Although acetylation of these particular lysine residues has not yet been detected in cross-linked peptide relationships, the fact that these sites of acetylation are known to be important for stabilization of protein interactions demonstrates that *in vivo* cross-linking methods can in some cases, be used to identify interactions topologies in these same critical regions. The relationships between sites of post-translational modifications that were

identified in cross-linked peptides from other proteins are discussed in more detail below.

Another example of a cross-linked homodimer is the mitochondrial enzyme glutamate dehydrogenase (GDH). GDH exists as a homo-hexamer and catalyzes the conversion of glutamate into α -ketoglutarate and ammonia. PIR data allowed identification of the peptide FGK⁴⁷⁹HGGTIP-IVPTAEFQDR as an unambiguous cross-linked homodimer [supplemental Table S2](#). In a similar situation to the cases discussed above, the cross-linked lysine residue (K479) is also known to be a site of acetylation (46). GDH has been identified as an *in vivo* target of the sirtuin SIRT3, although the functional significance of GDH acetylation remains unclear (49). The cross-linked site exists near a tri-molecular interface at the tip of the antenna domain [supplemental Fig. S7](#). The antenna domain is not found in bacterial, plant or fungal GDH and is thought to play an important role in allosteric regulation of GDH (50).

Extensive cross-linking of histones—From the high confidence set of 368 cross-linked pairs, 162 (44%) were intra- or interprotein links between histone proteins. Histones are the chief protein components of chromatin, forming a bead like nucleosome core complex around which DNA is coiled. There are five major classes of histones including the core histones H2A, H2B, H3, and H4, and the linker histones H1/H5. Experiments reported here resulted in identification of cross-links in and between each of these classes of histones. A nucleosome particle is comprised of an octameric complex containing two copies of each of the four core histone proteins around which 147 base pairs of DNA is wrapped. The structure of the core histones is highly conserved consisting of a helix-turn-helix-turn-helix motif from which long tails extend. The histone tails are highly disordered in structure and enriched in Lys and Arg residues making them particularly basic. The tails play a particularly important role in epigenetic regulation of chromatin serving as a scaffold for a host of post-translational modifications including methylation, acetylation, phosphorylation, and others. It has been suggested that combinations of these modifications may alter histone topology and interactions serving to regulate chromatin function in a complex chemical language known as the “histone code” (51). The alkaline property of histones may in part explain why such a large number of cross-links in and among these proteins is present in these data.

Mapping the observed histone cross-links onto the human chromatin x-ray crystal structure (PDB 3AFA) (52), enables reconstruction of the assembly of the octamer complex from information contained in the cross-linked sites at multiple levels (intraprotein, homodimer, and interprotein) (Fig. 4). Importantly the information provided by these cross-links sheds light on the structure and orientation of the nucleosome complexes as present *in vivo*. Fig. 4A illustrates the cross-linker reactive residues observed for each of the four types of core histone proteins. Disordered N-terminal and C-terminal re-

gions not included in the crystal structure were drawn in manually to illustrate the multiple cross-linked sites observed on the histone tails. Interprotein cross-links including unambiguous homodimer cross-links between H3 and H4, and H2A and H2B are displayed on the tetramer structures in Fig. 4B; however, the N-terminal tails are excluded here for clarity. Finally cross-links between the H3-H4 subunits and the H2A-H2B subunits are displayed with the full chromatin structure in Fig. 4C.

Cross-links Containing Post-translational Modifications—Histone H31 was the most heavily post-translationally modified protein detected in this study. In total, 13 unique post-translational modification sites on histone H31 were identified in cross-linked peptide pairs from human cells. These included the acetylation sites H3K14ac, H3K23ac, and H3K27ac, the mono-methylation sites at H3K9me, H3K27me, H3K36me and H3K37me, di-methylation sites H3K9me₂, H3K27me₂, H3K36me₂, and H3K37me₂, and trimethylation sites H3K9me₃, and H3K27me₃. We have mapped the modifications observed at each site along with the observed cross-links onto the sequence for histone H3 in Fig. 5. It should be noted that except for the cases of unambiguous homodimer cross-links, these data are not able to conclusively distinguish intramolecular from intermolecular linkages in histone tails. Regardless, these results provide interesting insight into the topology of histone H3 and how it is altered with varying post-translational modification. For instance, cross-links between residues K4-K23, K18-K18, and K18-K23 are only observed with unmodified peptides. In contrast cross-links between residues K4-K36, K4-K37, K14-K14, K14-K23, K14-K36, K14-K37, K14-K56, K23-K36, K23-K37, and K37-K56 of histone H3 are only observed with mono-, di-, or tri-methylation present on a site on one of the cross-linked peptides. Linkages unique to acetylation modifications include K4-K18 and K9-K18. Overlap of the histone H3 intralinks with the presence of post-translational modifications is illustrated by the Venn diagram in Fig. 5. Interestingly, linkages between the end of the N-terminal tail (K4) and the base of the tail (K36 and K37) are only observed when there is a di- or tri-methylation at K27. Similarly cross-links between K14 and K23, K36, K37, and K53 are only observed with mono-, di-, or tri- methylation at H3K9. Various degrees of methylation at H3K9 are known to each have distinct effects over chromatin structure and activation or repression of specific genes (53). Cross-links were observed between H3K14 and K60 of N-actetyltransferase 10 (NAT10), a protein known to acetylate histones and stimulate telomerase activity (54), when either H3K9me₂ or H3K9me₃ were present suggesting these modifications could be important for this interaction. The interconversion between the different states of methylation at H3K9 is regulated by a diverse set of methyltransferases and demethylases (55). Partial chromatographic resolution of cross-linked peptides between residues 14 and 18 containing various degrees of methyl-

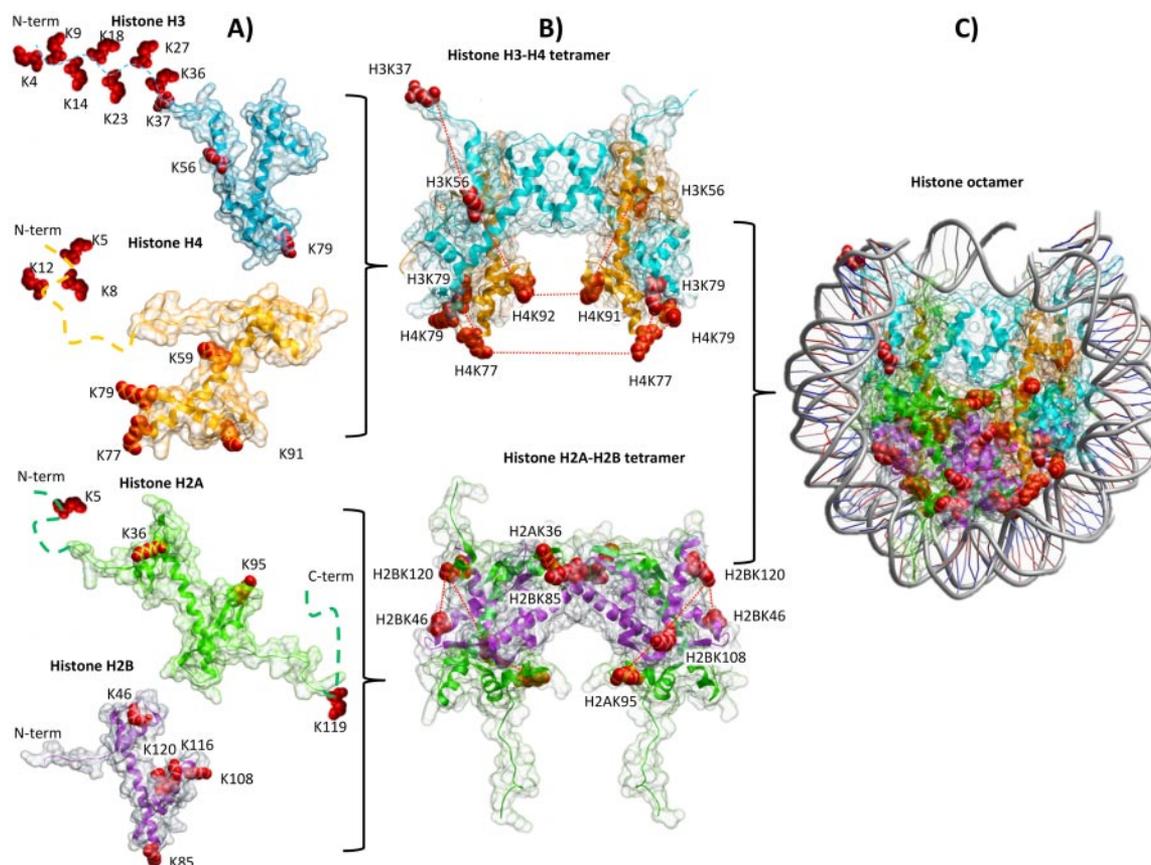


FIG. 4. Cross-links mapped onto structure of nucleosome (PDB = 3AFA). A, Individual monomer structures of the four core histone proteins with cross-linker reactive lysine residues highlighted in red space filling display. N-terminal and C-terminal tails not present in the crystal structure were drawn manually (indicated by dashed lines) to illustrate cross-linked sites in these highly disordered regions. B, Tetramer structures for H3₂-H4₂ and H2A₂-H2B₂ with intraprotein and interprotein cross-links displayed as red dashes. C, Complete nucleosome particle including 137bp DNA wrapped around histone octamer complex with cross-links displayed.

tion at H3K9 is shown in Fig. 5. As expected, with increasing degrees of methylation the retention time is lengthened on average. In addition, the integrated chromatographic peak areas of each of the modified forms is different, with dimethylation at H3K9 having the largest area and the unmodified form having the smallest area, though it should be noted peak areas may not be directly comparable across modification states because of differing ionization efficiencies. However the ability to chromatographically separate cross-linked peptides with differing modification states opens the possibility to quantify changes to various forms across different biological states employing stable isotope labeling techniques. The information obtained by such measurements would be distinct from global levels of modification at a specific site because of the topological information contained in the cross-linked peptide pair.

For the case of histone H4, acetylation modification was observed at H4K16ac. The intraprotein cross-link between K5-K12 was observed in the presence and absence of H4K16ac. Similarly a cross-link between H4K12 and H2BK108 was observed in the presence and absence of H4K16ac. Acetylation at H4K16ac has been shown to inhibit

formation higher order chromatin structure contributing to de-condensation of chromatin fibers (56). Furthermore the acetylation state of H4K16 has been shown to regulate interactions between various forms of chromatin and interacting proteins including Sir3, ISWI, and Bdf1 (57). We also identified two post-translationally modified sites of elongation factor 1-alpha (EF1A1) in cross-linked relationships including trimethylation on K35 and dimethylation on K54. Both of these modifications have been previously observed in EF1A1 isolated from rabbit reticulocytes (58). Although the biological roles of methylation on these two sites of EF1A1 have not been characterized, Lamberti *et al.* propose that these modifications increase the enzymatic activity of EF1A1 (59). EF1A1 is a core component of the protein synthesis machinery promoting the GTP-dependent binding of aminoacyl-tRNA to the A-site of ribosomes during protein biosynthesis however has additional roles in cell signaling and apoptosis pathways (59).

As demonstrated by these exciting results, it is now possible to directly monitor the topological effects of post-translational modifications at discrete sites in proteins using *in vivo* cross-linking with mass spectrometry. This opens the door to

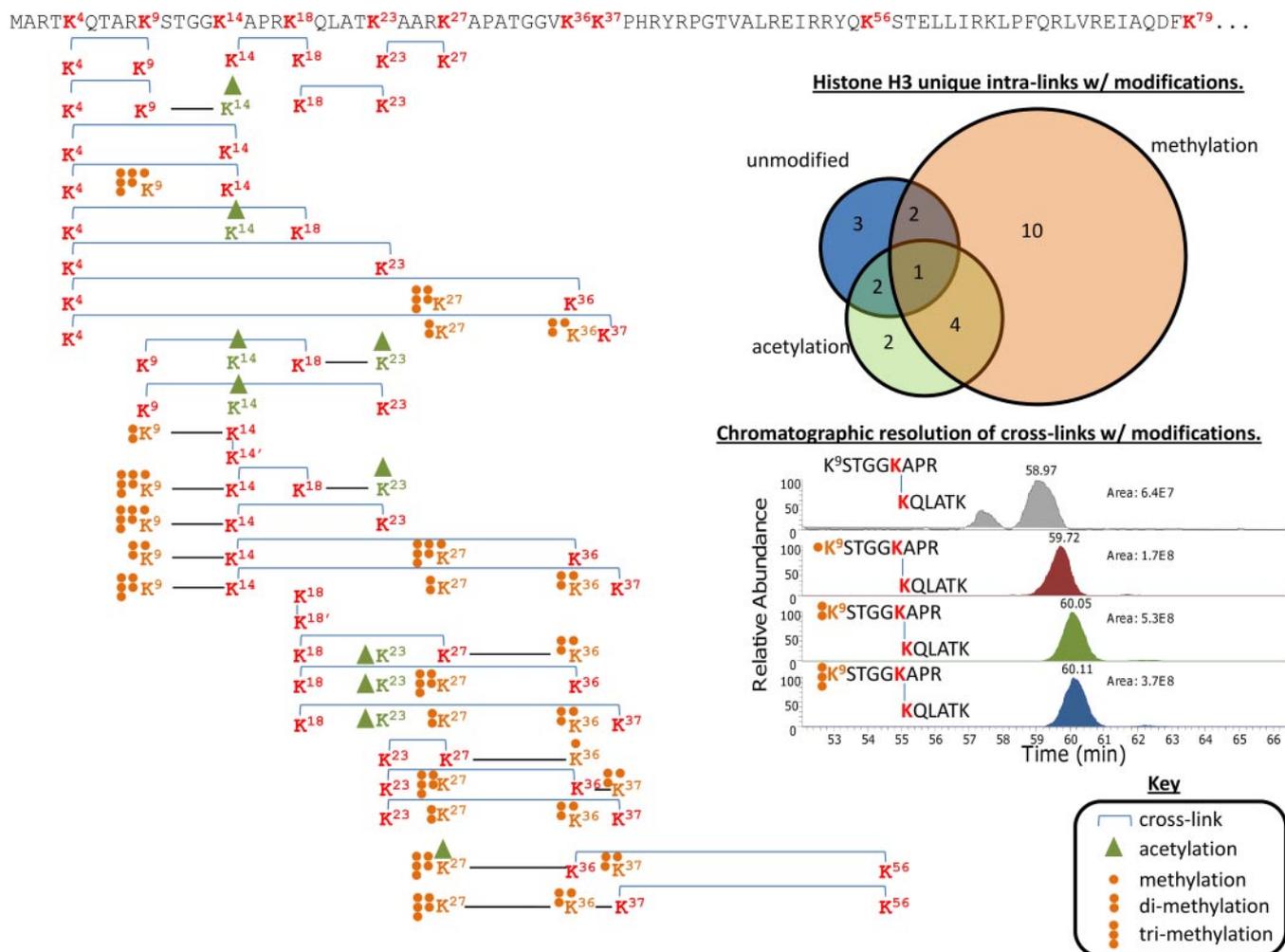


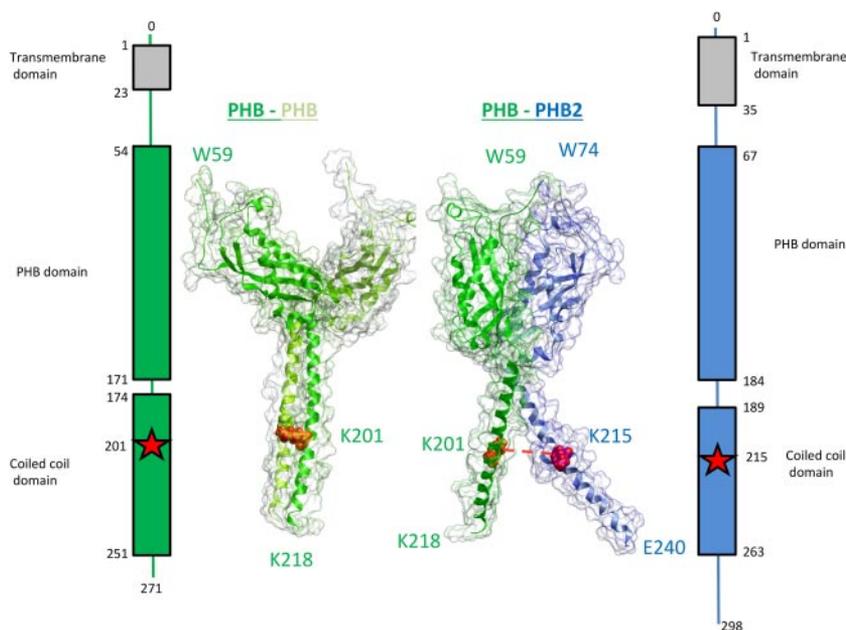
FIG. 5. **Cross-link map of histone H3 including post-translational modifications.** Sequence of histone H3 from residues 0–79 with cross-linked sites highlighted in bold red with residue numbers in superscript. Mapped cross-links are shown below and include post-translational modifications as indicated in the key. Venn diagram illustrates overlap of cross-links observed between unmodified, acetylated, or methylated (mono-, di-, tri- methylation grouped together). Extracted ion chromatographs are included for cross-linked peptides linking K14-K18 illustrating chromatographic resolution of various modified forms of this cross-linked peptide pair.

many future proteomics experiments in which the effects of varying levels and types of post-translational modifications across differing biological states can be directly linked to changes in protein topology and interactions.

New Insights into Interaction Topologies—These cross-linking results provide new insight into protein interactions as they exist in the cell. This can be in the form of novel interacting partners or new topological information on known protein-protein interactions for which no previous structural information exists. One such example is the known interacting partners prohibitin (PHB) and prohibitin-2 (PHB2). Prohibitins are highly conserved, ubiquitous, and pleiotropic proteins implicated in a diversity of biological processes including proliferation, regulation of transcription, apoptosis and cellular senescence. Evidence from yeast suggests prohibitins primarily localize to the inner mitochondrial membrane where PHB and PHB2 (aka BAP32 and BAP37) assemble into a ring

shaped complex of ~1.2–1.4 MDa consisting of ~14 PHB-PHB2 dimers (60). The stabilities of PHB and PHB2 are also linked as they are readily degraded in the absence of their respective partner (61). In addition to their role in mitochondrial function, evidence also indicates prohibitins localize to the nuclear and the plasma membranes where they function in transcriptional regulation and signal transduction (61). Prohibitins are also emerging as potential therapeutic targets because of evidence implicating them in human health disorders including HIV, cancer, inflammatory disorders, diabetes, and obesity (62, 63). Therefore there is much interest in understanding the molecular mechanisms by which prohibitins are able to carry out their diverse functions. Membrane proteins such as the prohibitins are notoriously difficult to study with structural techniques such as NMR and x-ray crystallography and unfortunately, structural details on prohibitins are scarce. Back *et al.* employed a combination of structural modeling

FIG. 6. Model structures for PHB-PHB homodimer and PHB-PHB2 dimer generated through homology modeling and molecular docking using distance constraints from cross-linked residues. The cross-linked sites PHB K201 and PHB2 K215 are located in the C-terminal domain thought to be important for stabilizing this interaction.



and cross-linking on electrophoretically purified prohibitin complex from yeast mitochondria to derive a model for the PHB-PHB2 building block and a superstructure ring complex (7).

From cross-linking in HeLa cells we identified a cross-link between K201 of PHB and K215 of PHB2. Importantly these sites exist within predicted coiled-coil domains of PHB and PHB2 thought to be important for interaction between prohibitin subunits (64). Interestingly, the site of *in vivo* cross-linking between PHB and PHB2 in human cells reported here is conserved *in vitro* in purified yeast complexes where K204 was identified as cross-linked to K233 of PHB2 (supplemental Fig. S8) (7). To construct a molecular model for the PHB-PHB2 dimer we first obtained homology models for PHB (residues 59–218, with 99.9% confidence) and PHB2 (residues 73 to 239, with 100% confidence) monomers using the protein structure prediction software Phyre2 (26). Both models were constructed using the crystal structure of a core domain of stomatin from *Pyrococcus horikoshii* (PDB = 3BK6) (sequence alignment provided in supplemental Fig. S8). The monomers were docked using PatchDock (27) using distance constraints derived from the cross-linked residues identified here. The top scoring PHB-PHB2 dimer and PHB-PHB homodimer models from PatchDock are shown in Fig. 6 with the PDB files included in supplemental Data Set S2. Although slight differences exist between the homodimer and heterodimer models, the alpha helical C-terminal domains in both of these models are interacting when cross-linking restraints are applied. For comparison, dimer models for these two complexes were generated without cross-linking distance constraint information and are shown in supplemental Fig. S9. It can be seen that without applying the information from the cross-

linked sites the resulting dimer models are quite different with the C-terminal domains on opposite sides of the complex. Despite the previous lack of evidence for any PHB homo-oligomers (7, 60), our results also provide the first direct evidence for a prohibitin homodimer with an unambiguous peptide homodimer cross-link observed between K201 of PHB. Taken together, our results and those of previous studies suggest these lysine sites to be important for homo- and hetero- interactions in human prohibitin. Knowledge of this interaction topology could be of potential use in future development of therapeutics targeting PHB.

Serving as another example of new protein interaction topology revealed in these data is the cross-link between K591 of stabilin-1 (STAB1) and K563 of ribophorin-1 (RPN1). There are no existing structures for either of these proteins in the PDB. RPN1 is an essential component of the N-oligosaccharyl transferase (OST) complex responsible for the transfer of oligosaccharides from dolichol to N-X-(S/T) motifs on nascent membrane proteins. RPN1 has been shown to transiently associate with a subset of newly synthesized membrane proteins immediately upon leaving the Sec61 translocon (65). Results from *in vitro* cross-linking experiments have suggested RPN1 serves to bind and deliver substrate proteins to the catalytic core of the OST (66). However, there is no existing evidence for interaction between RPN1 and STAB1 and these proteins are separated by two nodes in the IntAct database. STAB1 is a transmembrane receptor glycoprotein with ascribed functions in endocytosis, angiogenesis, inflammation, cell adhesion, and cell-cell interactions among others (67). STAB1 contains 7 fasciclin (FAS), 16 epidermal growth factor (EGF)-like, and 2 laminin-type EGF-like domains as well as a C-type lectin-like hyaluronan-binding Link module (68). The site of cross-linking (K591 of

RPN1, K563 of STAB1) links a predicted cytoplasmic domain on RPN1 (residues 457–606) to the second extracellular FAS domain in STAB1 (residues 505–640). This FAS domain also contains a single N-glycosylation motif (NIS, residues 605–607). These results identify STAB1 as a potential novel substrate of RPN1.

Concluding Remarks—Undoubtedly many challenges remain to allow the cross-linking field to achieve its full potential for large-scale *in vivo* topological studies. Although the goal of the present manuscript is to demonstrate utility for protein interaction topological measurements in human cells, future studies will likely target quantitative measurements and potentially allow visualization of dynamic changes in protein interactions and topologies. The ultimate goal of complete coverage of the interactome will continue to demand advanced method and technology development for cross-linking studies. The primary advancements demonstrated here include the application of cross-linking to living human cells, mapping protein topologies and interactions in their *in vivo* state, as well as the ability to detect cross-linked peptide pairs containing post-translational modifications. Chemical cross-linking provides unique information on protein structural changes induced by post-translational modifications in cells, which may ultimately drive changes in biological function. This represents an exciting and critical area of systems biology research for which, no other large-scale technology currently exists. With further advancement, cross-linking technologies will evolve to fulfill this need and link disparate aspects of post-translational modifications, protein topologies and protein interactions.

Acknowledgments—We thank members of the Bruce Laboratory and UWPR for helpful input and discussions.

* This work was supported by the National Institutes of Health through grant numbers: 5R01RR023334, 1R01GM097112, 5R01GM086688, 7S10RR025107 and 1R01HL110879.

☐ This article contains [supplemental Data Sets S1 and S2, Figs. S1 to S7, and Tables S1 to S3](#).

§ To whom correspondence should be addressed: University of Washington, 850 Republican St., Seattle, WA 98109. Tel.: 206.543.0220; Fax: 206.616.0008; E-mail: jimbruce@uw.edu.

REFERENCES

- Hopwood, D. (1969) Fixatives and fixation: a review. *Histochem. J.* **1**, 323–360
- Dutton, A., Adams, M., and Singer, S. J. (1966) Bifunctional imidoesters as cross-linking reagents. *Biochem. Biophys. Res. Commun.* **23**, 730–739
- Kluger, R., and Alagic, A. (2004) Chemical cross-linking and protein-protein interactions—a review with illustrative protocols. *Bioorg. Chem.* **32**, 451–472
- Wold, F. (1961) The reaction of bovine serum albumin with the bifunctional reagent p,p'-difluoro-m,m'-dinitro-diphenyl-sulfone. *J. Biol. Chem.* **236**, 106–111
- Leitner, A., Walzthoeni, T., Kahraman, A., Herzog, F., Rinner, O., Beck, M., and Aebersold, R. (2010) Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics. *Mol. Cell. Proteomics* **9**, 1634–1649
- Chu, F., Shan, S. O., Moustakas, D. T., Alber, F., Egea, P. F., Stroud, R. M., Walter, P., and Burlingame, A. L. (2004) Unraveling the interface of signal recognition particle and its receptor by using chemical cross-linking and tandem mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 16454–16459
- Back, J. W., de Jong, L., Muijsers, A. O., and de Koster, C. G. (2003) Chemical cross-linking and mass spectrometry for protein structural modeling. *J. Mol. Biol.* **331**, 303–313
- Serpa, J. J., Parker, C. E., Petrotschenko, E. V., Han, J., Pan, J., and Borchers, C. H. (2012) Mass spectrometry-based structural proteomics. *Eur. J. Mass Spectr.* **18**, 251–267
- Fields, S., and Song, O. (1989) A novel genetic system to detect protein-protein interactions. *Nature* **340**, 245–246
- Bauer, A., and Kuster, B. (2003) Affinity purification-mass spectrometry. Powerful tools for the characterization of protein complexes. *Eur. J. Biochem. FEBS* **270**, 570–578
- Herzog, F., Kahraman, A., Boehringer, D., Mak, R., Bracher, A., Walzthoeni, T., Leitner, A., Beck, M., Hartl, F. U., Ban, N., Malmström, L., and Aebersold, R. (2012) Structural Probing of a Protein Phosphatase 2A Network by Chemical Cross-Linking and Mass Spectrometry. *Science* **337**, 1348–1352
- Lauber, M. A., and Reilly, J. P. (2011) Structural analysis of a prokaryotic ribosome using a novel amidinating cross-linker and mass spectrometry. *J. Proteome Res.* **10**, 3604–3616
- Chavez, J. D., Cilia, M., Weisbrod, C. R., Ju, H. J., Eng, J. K., Gray, S. M., and Bruce, J. E. (2012) Cross-linking measurements of the Potato leafroll virus reveal protein interaction topologies required for virion stability, aphid transmission, and virus-plant interactions. *J. Proteome Res.* **11**, 2968–2981
- Zhang, H., Tang, X., Munske, G. R., Zakharova, N., Yang, L., Zheng, C., Wolff, M. A., Tolic, N., Anderson, G. A., Shi, L., Marshall, M. J., Fredrickson, J. K., and Bruce, J. E. (2008) In vivo identification of the outer membrane protein OmcA-MtrC interaction network in *Shewanella oneidensis* MR-1 cells using novel hydrophobic chemical cross-linkers. *J. Proteome Res.* **7**, 1712–1720
- Rinner, O., Seebacher, J., Walzthoeni, T., Mueller, L. N., Beck, M., Schmidt, A., Mueller, M., and Aebersold, R. (2008) Identification of cross-linked peptides from large sequence databases. *Nat. Methods* **5**, 315–318
- 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., and Dong, M. Q. (2012) Identification of cross-linked peptides from complex samples. *Nat. Methods* **9**, 904–906
- Zheng, C., Yang, L., Hoopmann, M. R., Eng, J. K., Tang, X., Weisbrod, C. R., and Bruce, J. E. (2011) Cross-linking measurements of in vivo protein complex topologies. *Mol. Cell. Proteomics* **10**, M110 006841
- Yang, L., Zheng, C., Weisbrod, C. R., Tang, X., Munske, G. R., Hoopmann, M. R., Eng, J. K., and Bruce, J. E. (2012) In vivo application of photo-cleavable protein interaction reporter technology. *J. Proteome Res.* **11**, 1027–1041
- Weisbrod, C. R., Chavez, J. D., Eng, J. K., Yang, L., Zheng, C., and Bruce, J. E. (2012) In vivo protein interaction network identified with novel chemical cross-linking technology. *J. Proteome Res.* Available at: 10.1021/pr3011638
- Zheng, C., Weisbrod, C. R., Chavez, J. D., Eng, J. K., Sharma, V., Wu, X., and Bruce, J. E. (2012) CrossLink-DB: database and software tools for storing and visualizing protein interaction topology data. *J. Proteome Res.* Available at: 10.1021/pr301162j
- Bruce, J. E. (2012) In vivo protein complex topologies: sights through a cross-linking lens. *Proteomics* **12**, 1565–1575
- Vidal, M., Cusick, M. E., and Barabasi, A. L. Interactome networks and human disease. *Cell* **144**, 986–998
- Bich, C., Maedler, S., Chiesa, K., DeGiacomo, F., Bogliotti, N., and Zenobi, R. (2010) Reactivity and applications of new amine reactive cross-linkers for mass spectrometric detection of protein-protein complexes. *Anal. Chem.* **82**, 172–179
- Anderson, G. A., Tolic, N., Tang, X., Zheng, C., and Bruce, J. E. (2007) Informatics strategies for large-scale novel cross-linking analysis. *J. Proteome Res.* **6**, 3412–3421
- Peng, K., Radivojac, P., Vucetic, S., Dunker, A. K., and Obradovic, Z. (2006) Length-dependent prediction of protein intrinsic disorder. *BMC Bioinformatics* **7**, 208
- Kelley, L. A., and Sternberg, M. J. (2009) Protein structure prediction on the

- Web: a case study using the Phyre server. *Nat. Protocols* **4**, 363–371
27. Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., and Wolfson, H. J. (2005) PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res.* **33**, W363–367
 28. Tang, X., Yi, W., Munske, G. R., Adhikari, D. P., Zakharova, N. L., and Bruce, J. E. (2007) Profiling the membrane proteome of *Shewanella oneidensis* MR-1 with new affinity labeling probes. *J. Proteome Res.* **6**, 724–734
 29. Zhang, H., Tang, X., Munske, G. R., Tolic, N., Anderson, G. A., and Bruce, J. E. (2009) Identification of protein-protein interactions and topologies in living cells with chemical cross-linking and mass spectrometry. *Mol. Cell. Proteomics* **8**, 409–420
 30. Oeffinger, M. (2012) Two steps forward—one step back: advances in affinity purification mass spectrometry of macromolecular complexes. *Proteomics* **12**, 1591–1608
 31. Kerrien, S., Aranda, B., Breuza, L., Bridge, A., Broackes-Carter, F., Chen, C., Duesbury, M., Dumousseau, M., Feuermann, M., Hinz, U., Jandrasits, C., Jimenez, R. C., Khadake, J., Mahadevan, U., Masson, P., Pedruzzi, I., Pfeifferberger, E., Porras, P., Raghunath, A., Roehert, B., Orchard, S., and Hermjakob, H. (2012) The IntAct molecular interaction database in 2012. *Nucleic Acids Res.* **40**, D841–846
 32. Schwikowski, B., Uetz, P., and Fields, S. (2000) A network of protein-protein interactions in yeast. *Nat. Biotechnol.* **18**, 1257–1261
 33. Roca, J., and Wang, J. C. (1994) DNA transport by a type II DNA topoisomerase: evidence in favor of a two-gate mechanism. *Cell* **77**, 609–616
 34. Kahraman, A., Malmstrom, L., and Aebersold, R. (2011) Xwalk: computing and visualizing distances in cross-linking experiments. *Bioinformatics* **27**, 2163–2164
 35. Ispolatov, I., Yuryev, A., Mazo, I., and Maslov, S. (2005) Binding properties and evolution of homodimers in protein-protein interaction networks. *Nucleic Acids Res.* **33**, 3629–3635
 36. Kuriyan, J., and Eisenberg, D. (2007) The origin of protein interactions and allostery in colocalization. *Nature* **450**, 983–990
 37. Maiolica, A., Cittaro, D., Borsotti, D., Sennels, L., Ciferri, C., Tarricone, C., Musacchio, A., and Rappsilber, J. (2007) Structural analysis of multiprotein complexes by cross-linking, mass spectrometry, and database searching. *Mol. Cell. Proteomics* **6**, 2200–2211
 38. Whitesell, L., and Lindquist, S. L. (2005) HSP90 and the chaperoning of cancer. *Nat. Rev. Cancer* **5**, 761–772
 39. Ali, M. M., Roe, S. M., Vaughan, C. K., Meyer, P., Panaretou, B., Piper, P. W., Prodromou, C., and Pearl, L. H. (2006) Crystal structure of an Hsp90-nucleotide-p23/Sba1 closed chaperone complex. *Nature* **440**, 1013–1017
 40. Pearl, L. H., and Prodromou, C. (2006) Structure and mechanism of the Hsp90 molecular chaperone machinery. *Annu. Rev. Biochem.* **75**, 271–294
 41. Li, Y., Zhang, T., Schwartz, S. J., and Sun, D. (2009) New developments in Hsp90 inhibitors as anti-cancer therapeutics: mechanisms, clinical perspective and more potential. *Drug Resistance Updates* **12**, 17–27
 42. Chen, B., Piel, W. H., Gui, L., Bruford, E., and Monteiro, A. (2005) The HSP90 family of genes in the human genome: insights into their divergence and evolution. *Genomics* **86**, 627–637
 43. Minami, Y., Kawasaki, H., Miyata, Y., Suzuki, K., and Yahara, I. (1991) Analysis of native forms and isoform compositions of the mouse 90-kDa heat shock protein, HSP90. *J. Biol. Chem.* **266**, 10099–10103
 44. Nemoto, T., and Sato, N. (1998) Oligomeric forms of the 90-kDa heat shock protein. *Biochem. J.* **330** (Pt 2), 989–995
 45. Perdew, G. H., Hord, N., Hollenback, C. E., and Welsh, M. J. (1993) Localization and characterization of the 86- and 84-kDa heat shock proteins in Hepa 1c1c7 cells. *Exp. Cell Res.* **209**, 350–356
 46. Choudhary, C., Kumar, C., Gnad, F., Nielsen, M. L., Rehman, M., Walther, T. C., Olsen, J. V., and Mann, M. (2009) Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* **325**, 834–840
 47. Kim, S. C., Sprung, R., Chen, Y., Xu, Y., Ball, H., Pei, J., Cheng, T., Kho, Y., Xiao, H., Xiao, L., Grishin, N. V., White, M., Yang, X. J., and Zhao, Y. (2006) Substrate and functional diversity of lysine acetylation revealed by a proteomics survey. *Mol. Cell* **23**, 607–618
 48. Scroggins, B. T., Robzyk, K., Wang, D., Marcu, M. G., Tsutsumi, S., Beebe, K., Cotter, R. J., Felts, S., Toft, D., Karnitz, L., Rosen, N., and Neckers, L. (2007) An acetylation site in the middle domain of Hsp90 regulates chaperone function. *Mol. Cell* **25**, 151–159
 49. Lombard, D. B., Alt, F. W., Cheng, H. L., Bunkenborg, J., Streeper, R. S., Mostoslavsky, R., Kim, J., Yancopoulos, G., Valenzuela, D., Murphy, A., Yang, Y., Chen, Y., Hirsche, M. D., Bronson, R. T., Haigis, M., Guarente, L. P., Farese, R. V., Jr., Weissman, S., Verdine, E., and Schwer, B. (2007) Mammalian Sir2 homolog SIRT3 regulates global mitochondrial lysine acetylation. *Mol. Cell. Biol.* **27**, 8807–8814
 50. Smith, T. J., Schmidt, T., Fang, J., Wu, J., Siuzdak, G., and Stanley, C. A. (2002) The structure of apo human glutamate dehydrogenase details subunit communication and allostery. *J. Mol. Biol.* **318**, 765–777
 51. Jenuwein, T., and Allis, C. D. (2001) Translating the histone code. *Science* **293**, 1074–1080
 52. Tachiwana, H., Kagawa, W., Osakabe, A., Kawaguchi, K., Shiga, T., Hayashi-Takanaka, Y., Kimura, H., and Kurumizaka, H. (2010) Structural basis of instability of the nucleosome containing a testis-specific histone variant, human H3T. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 10454–10459
 53. Snowden, A. W., Gregory, P. D., Case, C. C., and Pabo, C. O. (2002) Gene-specific targeting of H3K9 methylation is sufficient for initiating repression in vivo. *Current Biol.* **12**, 2159–2166
 54. Lv, J., Liu, H., Wang, Q., Tang, Z., Hou, L., and Zhang, B. (2003) Molecular cloning of a novel human gene encoding histone acetyltransferase-like protein involved in transcriptional activation of hTERT. *Biochem. Biophys. Res. Commun.* **311**, 506–513
 55. Black, J. C., and Whetstone, J. R. (2010) Chromatin landscape: methylation beyond transcription. *Epigenetics* **6**, 9–15
 56. Shogren-Knaak, M., Ishii, H., Sun, J. M., Pazin, M. J., Davie, J. R., and Peterson, C. L. (2006) Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* **311**, 844–847
 57. Kurdistani, S. K., Tavazoie, S., and Grunstein, M. (2004) Mapping global histone acetylation patterns to gene expression. *Cell* **117**, 721–733
 58. Dever, T. E., Costello, C. E., Owens, C. L., Rosenberry, T. L., and Merrick, W. C. (1989) Location of seven post-translational modifications in rabbit elongation factor 1 alpha including dimethyllysine, trimethyllysine, and glycerylphosphorylethanolamine. *J. Biol. Chem.* **264**, 20518–20525
 59. Lamberti, A., Caraglia, M., Longo, O., Marra, M., Abbruzzese, A., and Arcari, P. (2004) The translation elongation factor 1A in tumorigenesis, signal transduction and apoptosis: review article. *Amino Acids* **26**, 443–448
 60. Tatsuta, T., Model, K., and Langer, T. (2005) Formation of membrane-bound ring complexes by prohibitins in mitochondria. *Mol. Biol. Cell* **16**, 248–259
 61. Osman, C., Merkwirth, C., and Langer, T. (2009) Prohibitins and the functional compartmentalization of mitochondrial membranes. *J. Cell Sci.* **122**, 3823–3830
 62. Emerson, V., Holtkotte, D., Pfeiffer, T., Wang, I. H., Schnölzer, M., Kempf, T., and Bosch, V. (2010) Identification of the cellular prohibitin 1/prohibitin 2 heterodimer as an interaction partner of the C-terminal cytoplasmic domain of the HIV-1 glycoprotein. *J. Virol.* **84**, 1355–1365
 63. Theiss, A. L., and Sitaraman, S. V. (2011) The role and therapeutic potential of prohibitin in disease. *Biochim. Biophys. Acta* **1813**, 1137–1143
 64. Winter, A., Kämäräinen, O., and Hofmann, A. (2007) Molecular modeling of prohibitin domains. *Proteins* **68**, 353–362
 65. Wilson, C. M., Kraft, C., Duggan, C., Ismail, N., Crawshaw, S. G., and High, S. (2005) Ribophorin I associates with a subset of membrane proteins after their integration at the sec61 translocon. *chemistry. Biol. Chem.* **280**, 4195–4206
 66. Wilson, C. M., Roebuck, Q., and High, S. (2008) Ribophorin I regulates substrate delivery to the oligosaccharyltransferase core. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 9534–9539
 67. Kzhyskowska, J. (2010) Multifunctional receptor stabilin-1 in homeostasis and disease. *TheScientificWorldJournal* **10**, 2039–2053
 68. Politz, O., Gratchev, A., McCourt, P. A., Schledzewski, K., Guillot, P., Johansson, S., Svineng, G., Franke, P., Kannicht, C., Kzhyskowska, J., Longati, P., Velten, F. W., and Goerdt, S. (2002) Stabilin-1 and -2 constitute a novel family of fasciilin-like hyaluronan receptor homologues. *Biochem. J.* **362**, 155–164