Published in final edited form as:

Chem Res Toxicol. 2009 June ; 22(6): 1151–1162. doi:10.1021/tx900078y.

Proteomic Analysis of DNA-Protein Cross-Linking by Antitumor Nitrogen Mustards

Rachel L. Loeber[†], Erin D. Michaelson-Richie[†], Simona G. Codreanu[‡], Daniel C. Liebler[‡], Colin R. Campbell[§], and Natalia Y. Tretyakova^{†,*}

[†]Department of Medicinal Chemistry and the Masonic Cancer Center, University of Minnesota, Minneapolis, MN 55455

[‡]Department of Biochemistry, Vanderbilt University School of Medicine, Nashville, Tennessee 37232

[§]Department of Pharmacology, University of Minnesota, Minneapolis, MN 55455

Abstract

Nitrogen mustards are antitumor agents used clinically for the treatment of a variety of neoplastic conditions. The biological activity of these compounds is typically attributed to their ability to induce DNA-DNA cross-links. However, nitrogen mustards are able to produce a variety of other lesions, including DNA-protein cross-links (DPCs). DPCs induced by nitrogen mustards are not well characterized because of their structural complexity and the insufficient specificity and sensitivity of previously available experimental methodologies. In the present work, affinity capture methodology in combination with mass spectrometry-based proteomics was employed to identify mammalian proteins that form covalent cross-links to DNA in the presence of a simple nitrogen mustard, mechlorethamine. Following incubation of 5'-biotinylated DNA duplexes with nuclear protein extracts, DPCs were isolated by affinity capture on streptavidin beads, and the cross-linked proteins were identified by HPLC-ESI⁺-MS/MS of tryptic peptides. Mechlorethamine treatment resulted in the formation of DPCs with nuclear proteins involved in chromatin regulation, DNA replication and repair, cell cycle control, transcriptional regulation, and cell architecture. Western blot analysis was employed to confirm protein identification and to quantify the extent of drug-mediated cross-linking. Mass spectrometry of amino acid-nucleobase conjugates found in total proteolytic digests revealed that mechlorethamine-induced DPCs are formed *via* alkylation of the N7 position of guanine in duplex DNA and cysteine thiols within the proteins to give *N*-[2-[S-cysteinyl]ethyl]-*N*-[2-(guan-7-yl)ethyl]methylamine lesions. The results described herein suggest that cellular exposure to nitrogen mustards leads to cross-linking of a large spectrum of nuclear proteins to chromosomal DNA, potentially contributing to the cytotoxic and mutagenic effects of these drugs.

Introduction

DNA-protein cross-links (DPCs)¹ are bulky macromolecular conjugates that can form following the sequential alkylation of DNA and proteins by *bis*-electrophiles. DPCs can be induced by a variety of cytotoxic, mutagenic, and carcinogenic agents, including ionizing radiation (1), transition metals (2), endogenous and exogenous aldehydes (3), and

* To whom correspondence should be addressed: The Cancer Center, University of Minnesota, Mayo Mail Code 806, 420 Delaware St SE, Minneapolis, MN 55455, USA. ph: 612-626-3432 fax: 612-626-5135 email: E-mail: trety001@umn.edu.

Supporting Information Available: SDS-PAGE separation of nitrogen mustard-induced DPCs in CHO and HeLa nuclear extracts following biotin capture enrichment, and western blot analysis of DPCs involving histone H4 can be found in the Supporting Information. This material is available free of charge via the Internet at <http://pubs.acs.org>.

chemotherapeutic drugs (4,5). *In vivo* half-lives of DPCs can range from hours to days, depending on their chemical stability and their ability to be recognized by repair proteins (2). Because of the propensity of DPCs to persist within a cell, they potentially play a role in the antitumor activity of DNA-alkylating drugs used in cancer chemotherapy. Similarly, endogenously formed DPCs may contribute to spontaneous mutagenesis, cytotoxicity, and aging.

Perhaps because of their inherent complexity, DPCs have received less attention than other types of DNA damage, and their biological consequences are not well understood. As a result of their bulky nature and their ability to distort DNA structure, DPCs are expected to block the binding and progression of protein complexes involved in crucial cellular processes, including DNA replication, transcription, repair, recombination, and chromatin remodeling (6). If not repaired, DPCs are likely to contribute to the cytotoxic and mutagenic effects of common antitumor agents, e.g. alkylnitrosoureas, cisplatin, and nitrogen mustards, while endogenously generated lesions may play a role in cancer, aging, and neurodegenerative diseases.

Nitrogen mustards comprise a class of bifunctional alkylating agents used clinically in the treatment of a wide variety of cancers. Drugs of this class, e.g. mechlorethamine, chlorambucil, and melphalan, contain two N-(2-chloroethyl) groups which can react with two nucleophilic sites within DNA and proteins to form cross-linked lesions (Scheme 1). Historically, the antitumor activity of these drugs has been attributed to their ability to form DNA-DNA cross-links (8,9). However, biophysical studies have demonstrated that nitrogen mustards also induce DNA-protein lesions (10,11), although the identities of the relevant cellular proteins and the cross-link structures have not been determined.

Previous studies in our laboratory have shown that the human DNA repair protein, *O*⁶-alkylguanine DNA alkyltransferase (AGT), is readily cross-linked to DNA *in vitro* in the presence of two representative nitrogen mustards, mechlorethamine and chlorambucil (12). Within a cell, AGT is responsible for the repair of promutagenic *O*⁶-alkylguanine lesions in DNA *via* transfer of the *O*⁶-alkyl group from guanine in DNA to an active site cysteine residue of AGT, Cys¹⁴⁵ (13). We employed tandem mass spectrometry of tryptic peptides to map AGT-DNA cross-linking sites to two active site cysteine residues within the protein, Cys¹⁴⁵ and Cys¹⁵⁰. HPLC-ESI-MS/MS of total digests revealed that mechlorethamine-induced DPCs had the structure of *N*-[2-[*S*-cysteinyl]ethyl]-*N*-[2-(guan-7-yl)ethyl]methylamine (Cys-N7G-EMA). Western blot analysis revealed that AGT-DNA cross-linking was concentration-dependent and was not inhibited by the presence of other cellular proteins (12). As AGT is over-expressed in many tumor types (14-19), it is conceivable that drug-induced AGT-DNA cross-links may contribute to the cytotoxicity of antitumor nitrogen mustards.

The goal of the present work was to identify other mammalian nuclear proteins that form DPCs following exposure to mechlorethamine. An affinity capture methodology coupled with shotgun proteomics and western blot analysis was used to demonstrate that over 50 human proteins can become covalently cross-linked to DNA when human nuclear protein extracts are incubated with DNA duplexes in the presence of mechlorethamine. Our results further suggest

¹Abbreviations: DPC, DNA-protein cross-link; CHO, Chinese hamster ovary; AGT, *O*⁶-alkylguanine DNA alkyltransferase; PMSF, phenylmethanesulfonyl fluoride; DTT, dithiothreitol; PARP, poly(ADP-ribose)polymerase; Ref-1, DNA-(apurinic or apyrimidinic site) lyase; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; EF-1 α , elongation factor 1- α 1; Ku, ATP-dependent DNA helicase subunit 2; XRCC1, X-ray repair cross-complementary protein 1; Cys-N7G-EMA, *N*-[2-[*S*-cysteinyl]ethyl]-*N*-[2-(guan-7-yl)ethyl] methylamine; Lys-N7G-EMA, *N*-[2-[*N*-(lysyl)ethyl]-*N*-[2-(guan-7-yl)ethyl]methylamine; DMEM; Dulbecco's Modified Eagle's Medium; MS, mass spectrometry; HPLC-ESI⁺-MS/MS, high performance liquid chromatography-electrospray tandem mass spectrometry; SRM, selected reaction monitoring; FDR, false discovery rate; XRCC5, X-ray repair cross-complementary protein 5; SSBR, single-strand break repair; BER, base excision repair; NER, nucleotide excision repair.

that mechlorethamine-induced DPCs involve the N7 position of guanine in DNA and the thiolate side chains of the cysteine residues within proteins.

Materials and Methods

Chemicals and Reagents

Mechlorethamine hydrochloride, phenylmethanesulfonyl fluoride (PMSF), pepstatin, leupeptin, aprotinin, dithiothreitol (DTT), and iodoacetamide were purchased from Sigma-Aldrich (Milwaukee, WI). Trypsin Gold was obtained from Promega (Madison, WI) and carboxypeptidase Y and proteinase K were purchased from Worthington Biochemical Corporation (Lakewood, NJ). Primary polyclonal antibodies to mammalian actin, nucleolin, histone H4, poly (ADP-ribose) polymerase (PARP), DNA-(apurinic or apyrimidinic site) lyase (Ref-1), glyceraldehyde-3-phosphate dehydrogenase (GAPDH), and elongation factor 1- α 1 (EF-1 α 1) were obtained from Santa Cruz Biotechnology (Santa Cruz, CA). The primary polyclonal antibody to the ATP-dependent DNA helicase subunit 2 (Ku) and the monoclonal antibody against X-ray repair cross-complementary protein 1 (XRCC1) were purchased from Lab Vision/NeoMarkers (Fremont, CA). The primary monoclonal antibody to O^6 -alkylguanine DNA alkyltransferase (AGT) and the primary polyclonal antibody to histone H4 were obtained from Millipore (Temecula, CA). Alkaline phosphatase-conjugated anti-mouse and anti-rabbit IgG secondary antibodies were purchased from Sigma (St. Louis, MO). Synthetic DNA oligodeoxynucleotides were prepared at the University of Minnesota's BioMedical Genomics Center, *N*-[2-[*S*-cysteinyl]ethyl]-*N*-[2-(guan-7-yl)ethyl]methylamine (Cys-N7G-EMA) and *N*-[2-[*N*-(lysyl)ethyl]-*N*-[2-(guan-7-yl)ethyl]methylamine (Lys-N7G-EMA) were prepared as previously described (12).

Cell Culture

Chinese hamster ovary cells expressing human recombinant O^6 -alkylguanine DNA alkyltransferase (CHO-AGT) were generously provided by Dr. Anthony E. Pegg (Pennsylvania State University) and maintained as exponentially growing monolayer cultures in α -MEM supplemented with 9% fetal bovine serum and G-418 (Geneticin, 1 mg/mL) in a humidified incubator at 37°C with 5% CO₂ (20). Human cervical carcinoma (HeLa) cells were a gift from Dr. Jonathan Marchant (University of Minnesota). The cells were maintained as exponentially growing monolayer cultures in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 9% FBS in a humidified incubator at 37°C with 5% CO₂.

Preparation of Nuclear Extracts

Nuclear protein extracts were prepared as described previously (21). Briefly, $\sim 10^8$ cells were harvested, washed with phosphate-buffered saline, and suspended in hypotonic buffer (10 mM Tris-HCl - pH 7.4/10 mM MgCl₂/10 mM KCl/1 mM DTT). PMSF was added to a final concentration of 1 mM, and the cells were broken by 20 strokes in a Dounce homogenizer. The resulting mixture was centrifuged at 1,000g for 8 min, and the sedimented nuclei were resuspended in hypotonic buffer containing 350 mM NaCl. Following the addition of protease inhibitors (1 mg/mL pepstatin, 0.1 mg/mL leupeptin, 1.5 mg/mL aprotinin, and 100 μ M PMSF), the nuclei were incubated on ice for 1 h. The extracted nuclei were centrifuged at 160,000g at 4°C for 30 min, and the nuclear proteins were isolated in the clear supernatant. Extracts were dialyzed overnight at 4°C against 10 mM Tris-HCl (pH 7.4) using Slide-A-Lyzer dialysis cassettes from Pierce Biotechnology (3.5 kDa molecular weight cut-off, Rockford, IL). Protein concentrations were determined *via* a colorimetric assay (22).

DNA-Protein Cross-Linking and Biotin Capture

5'-Biotinylated double-stranded oligodeoxynucleotides (5'-GGA GCT CGT GGC CTA-3' (+) strand, 3.12 nmol) were incubated with CHO-AGT or HeLa nuclear extracts (500 µg total protein) in 10 mM Tris-HCl, pH 7.4, in the presence or in the absence of mechlorethamine (0–1000 µM, 1 mL total volume) at 37°C for 3 h. To isolate proteins covalently bound to DNA, streptavidin sepharose high performance beads (GE Healthcare, Piscataway, NJ) were used to capture the biotinylated oligodeoxynucleotide along with any DPCs. To remove non-covalently bound proteins, the beads were washed with 1% SDS, 4 M urea, and 1 M NaCl. Biotinylated DNA was released from the beads by heating in the presence of 4X SDS-PAGE loading buffer (90°C for 15 min). Cross-linked proteins were released from DNA by thermal hydrolysis (70°C for 1 h) and analyzed by HPLC-ESI⁺-MS/MS and western blot analysis as described below.

Protein Identification by Mass Spectrometry

Cross-linking reactions were performed as described above. Following elution from streptavidin beads, proteins were separated by SDS-PAGE. The gels were stained with SimplyBlue SafeStain (Invitrogen, Carlsbad, CA) to visualize the proteins. Gel lanes were cut into 5 slices encompassing the entire molecular weight range, then further into 1 mm pieces, followed by washing with 100 mM ammonium bicarbonate. Each fraction was subjected to reduction with DTT and alkylation with iodoacetamide as described previously (23). To initiate proteolytic digestion, gel pieces were dehydrated with acetonitrile, dried under vacuum, and reconstituted in 25 mM ammonium bicarbonate containing 10 µg (0.43 nmol) of proteomics grade trypsin. The samples were digested overnight at 37°C. Tryptic peptides were extracted from the gel pieces using 0.1% aqueous formic acid/60% acetonitrile, evaporated to dryness, and resuspended in 0.1% aqueous formic acid prior to mass spectrometric analysis.

All mass spectrometric analyses were performed on a Thermo LTQ linear ion trap MS system equipped with a Thermo Surveyor solvent delivery system and a microelectrospray source. Peptides were resolved on a 100 µm × 11 cm fused silica capillary column packed with 5 µm, 300 Å Jupiter C18 (Phenomenex, Torrance, CA) eluted at 0.6 µL/min with 0.1 formic acid in water (A) and 0.1% formic acid in acetonitrile (B). The column temperature was maintained at 25°C. Solvent composition was held at 2% B for the first 15 min, followed by a linear increase to 25% B over the next 35 min and further to 90% B over the next 15 min, where it was held constant for the remaining 10 min. MS/MS spectra of peptides were acquired using data-dependent scanning, in which a single full ESI⁺-MS scan (400–2000 *m/z*) was followed by four MS/MS scans. Spectra were recorded using dynamic exclusion of previously analyzed precursors for 60 s.

The “ScanSifter” algorithm v0.1, an in-house developed software, read tandem mass spectra stored as centroided peak lists from Thermo RAW files and transcribed them to DTA files. Spectra that contained fewer than 6 peaks or that had less than 20 measured TIC did not result in DTAs. If 90% of the intensity of a tandem mass spectrum appeared at a lower *m/z* than the precursor ion, a single precursor charge was assumed; otherwise, the spectrum was processed under both double and triple precursor charge assumptions. Proteins were identified using the SEQUEST v.27 algorithm (24,25) on a high speed, multiprocessor Linux cluster in the Advanced Computing Center for Research & Education at Vanderbilt University using the human subset of the IPI human protein database, version 331, created 7/20/07 or Chinese hamster database, created 2/14/2008. To estimate false discovery rates, each sequence of the database was reversed and concatenated to the database, for a total of 135,168 entries for the human database and a total of 1686 entries for the hamster database. The database search encompassed tryptic peptides with a maximum of 5 missed cleavage sites for enzyme search and with a maximum number of 10 internal cleavage sites. Cysteines were expected to undergo

carboxamidomethylation (+57 Da) and methionines were allowed to be oxidized (+16 Da). Mechlorethamine-induced alkylation at cysteine (hydrolyzed monoadduct: +102 Da, cross-link to guanine: +234 Da) were specified as dynamic modifications to identify spectra of adducted peptides. Precursor ions were required to fall within 1.25 m/z of the position expected from their average masses, and fragment ions were required to fall within 0.5 m/z of their monoisotopic positions. The database searches produced raw identifications in SQT format (26).

Peptide identification, filtering, and protein assembly were done with IDPicker software 42 which filtered raw peptide identifications to a target false discovery rate (FDR) of 5%. The peptide filtering employed reversed sequence database match information to determine thresholds that yielded an estimated 5% FDR for the identifications of each charge state by the formula $FDR = (2R)/(R+F)$, where R is the number of passing reversed peptide identifications and F is the number of passing forward (normal orientation) peptide identifications. The second round of filtering removed proteins supported by less than two distinct peptide identifications in the analyses. Indistinguishable proteins were recognized and grouped. Parsimony rules were applied to generate a minimal list of proteins that explained all of the peptides that passed our entry criteria (27). Furthermore, statistical analyses were performed to ensure that the levels of proteins captured from treated samples were significantly higher than those in untreated controls. As a result, any protein identified in the mechlorethamine-treated samples that displayed MS spectral counts comparable to the untreated controls was disregarded.

Western Blot Analysis of Identified Proteins

CHO and HeLa nuclear protein extracts were exposed to 0-1000 μ M mechlorethamine in the presence of the 5'-biotinylated oligonucleotide as described above. The resulting DPCs were isolated by biotin capture, and the cross-linked proteins were released by thermal hydrolysis and separated by SDS-PAGE (10 % or 12 % acrylamide concentration). Proteins were transferred to nitrocellulose membranes (Bio-Rad, Hercules, CA) and immediately blocked in Tris-buffered saline containing 5% bovine serum albumin. Following a 2 h incubation with the appropriate primary antibody at room temperature, the blots were washed 3 times with Tris-buffered saline and incubated with the corresponding alkaline phosphatase-conjugated secondary antibody (1-2 h at room temperature or 4°C overnight). Following washing with Tris-buffered saline to remove unbound antibodies, the blots were developed with SIGMA Fast BCIP/NBT (Bio-Rad, Hercules, CA). The developed blots were scanned as image files, and the optical densities of the bands in the resulting image files were quantified using the ImageJ software/shareware available free of charge from the NIH web site (www.ncbi.nlm.nih.gov). The extent of DNA-protein cross-linking for each protein was estimated by comparing the band intensities of cross-linked samples to the total amounts present in the nuclear extract.

HPLC-ESI⁺-MS/MS Analysis of Cys-N7G-EMA

Double-stranded oligodeoxynucleotides (5'-GGA GCT CGT GGC CTA-3' (+) strand, 9.36 nmol) were incubated with HeLa nuclear extracts (1.5 mg total protein) in 10 mM TRIS-HCl buffer (pH 7.4) in the presence or in the absence of mechlorethamine (0 or 1000 μ M, 1 mL total volume) for 3 h at 37°C. The samples were heated to 80°C for 30 min to release N7 alkylguanine lesions, including DPCs, from the DNA backbone, followed by proteolytic digestion of proteins to peptides with trypsin (20 μ g, 0.86 nmol in 25 mM ammonium bicarbonate, 37°C overnight). To achieve complete digestion of tryptic peptides to amino acids, the samples were dried under vacuum, reconstituted in 500 μ L water, followed by incubation with carboxypeptidase Y (10 μ g, 0.16 nmol) and proteinase K (10 μ g, 0.35 nmol) for 48 h at room temperature. The digested mixtures were purified by HPLC to selectively enrich for Cys-N7G-EMA using an Agilent Technologies HPLC system (1100 model) incorporating a variable

wavelength UV detector and a semi-micro UV cell. UV absorbance at 280 nm was monitored. A Supelcosil LC-18-DB column (4.6 mm × 250 mm, 5 μm) was eluted with 20 mM ammonium acetate, pH 4.9 (A) and acetonitrile (B) at a flow rate of 1 mL/min. The gradient program began at 0% B, followed by a linear increase to 24% B in 24 min, and further to 60% B in 6 min. Fractions corresponding to the elution time of Cys-N7G-EMA (8-10 min) were collected, dried in a speed-vac concentrator, and reconstituted in solvent A prior to HPLC-ESI⁺-MS/MS analysis.

A Waters nanoAQUITY UPLC system interfaced to a Thermo-Finnigan TSQ Quantum Ultra mass spectrometer was used for the detection of Cys-N7G-EMA in total proteolytic digests from mechlorethamine-treated HeLa cells. Chromatographic separation was achieved using a Phenomenex Synergi C18 column (250 mm × 0.5 mm, 4 μm) eluted with 15 mM ammonium acetate, pH 5.0 (A) and methanol (B) at a flow rate of 12 μL/min. When employing a linear gradient of 2-14% B over the course of 15 min, Cys-N7G-EMA eluted at ~ 13 min. Electrospray ionization was achieved at a spray voltage of 3.2 kV and a capillary temperature of 250°C. CID was performed with Ar as a collision gas (1.0 mTorr) and a collision energy of 18 V. The MS parameters were optimized for maximum response during infusion of a standard solution of Cys-N7G-EMA. Analyses were performed in the selected reaction monitoring (SRM) mode using the transition corresponding to the neutral loss of guanine from protonated molecules of the conjugate (m/z 356.1 → 205.1), which is the major fragment ion observed upon CID of Cys-N7G-EMA (12).

Preparation of HeLa Whole Cell Extract

Cultures of HeLa cells (~ 5 × 10⁶ cells) were harvested, washed with phosphate-buffered saline, and suspended in 2 mL cell lysis buffer containing 20 mM Tris – pH 7.8/50 mM KCl/0.2% NP-40/1 mM PMSF/1 μg/mL leupeptin/1 μg/mL pepstatin/5% glycerol. The cell suspension was sonicated three times (15 sec/each) on ice, and the resulting solution was then centrifuged at 10,000 g for 15 min at 4°C. The protein concentration of supernatant was determined *via* a colorimetric assay (22).

Western Blot Analysis of DPCs Involving Histone H4: Whole Cell Lysates

5'-Biotinylated double-stranded oligodeoxynucleotides (5'-GGA GCT CGT GGC CTA-3' (+) strand, 3.12 nmol) were combined with whole cell extract from HeLa cells (500 μg total protein) in 10 mM Tris-HCl – pH 7.4 in the absence or in the presence of mechlorethamine (0, 500, and 1000 μM – 1 mL total volume). The reaction mixtures were incubated at 37°C for 3 h to induce cross-linking. Proteins covalently bound to the biotinylated oligodeoxynucleotides were captured on streptavidin beads. To remove any non-covalently bound proteins, the beads were washed with 1% SDS, 4 M urea, and 1 M NaCl. Biotinylated DNA along with any cross-linked proteins was released from the beads by heating in the presence of 4X SDS-PAGE loading buffer (90°C for 15 min). Protein mixtures were separated by 15% SDS-PAGE prior to western blot analysis. Human recombinant histone H4 (Millipore, Billerica, MA) served as a positive control.

Results

Mass Spectrometric Identification of Cross-Linked Proteins

An affinity-based approach developed in our laboratory was employed to capture proteins that form covalent cross-links to DNA in the presence of mechlorethamine (Scheme 2). Nuclear protein extracts from transgenic Chinese hamster ovary cells expressing human recombinant AGT protein (CHO-AGT) or from human cervical carcinoma (HeLa) cells were incubated with synthetic DNA duplexes containing a 5'-biotin tag [5'-GGAGCTGGTGGCGTAGGC-3', (+) strand] in the presence or in the absence of mechlorethamine. Following capture of biotinylated

DNA on streptavidin beads and stringent washing steps to remove any non-covalently bound proteins, DPCs were eluted by heating in SDS-containing gel loading buffer. Because alkylation of DNA by mechlorethamine induces thermally labile N7-alkylguanine adducts (28), these conditions also release the protein component of the DPC from the DNA backbone in the form of protein-guanine conjugates (Scheme 2). SDS-PAGE analysis revealed little to no protein was in control samples lacking mechlorethamine, indicating that our washing procedures effectively remove any non-covalently bound proteins (see lane 3 in Figures 1A and 1B). In contrast, a concentration-dependent proteins capture indicative of DPC formation was observed in samples treated with 250, 500, 750, or 1000 μ M mechlorethamine (Figure 1 lanes 4-8).

To identify the proteins participating in cross-linking, nuclear extracts from CHO-AGT and HeLa cells (in triplicate) were treated with 0 (control) or 500 μ M of mechlorethamine in the presence of biotinylated DNA. Following biotin capture (Scheme 2), proteins were separated by SDS-PAGE (Supplements S-1 and S-2), and the gel bands encompassing the entire molecular weight range were excised for mass spectrometric analysis. The cross-linked proteins were subjected to in-gel tryptic digestion (23), and the resulting peptides were extracted from the gel and analyzed by HPLC-ESI⁺-MS/MS to determine the identities of cross-linked proteins. As shown for representative peptides in Figure 2, MS/MS spectra of tryptic fragments generated characteristic **b** and **y** series ions which were used to determine amino acid sequence of the proteins. Parsimony analysis (27) of the spectral data resulted in the identification of 15 CHO and 53 HeLa nuclear proteins that participated in DPC formation in the presence of mechlorethamine (Tables 1 and 2). We found that mechlorethamine-induced DPCs involved proteins of various cellular function, including architectural/structural proteins, chromatin regulators, proteins participating in cellular homeostasis/cell cycle, transcription regulators/RNA splicing, and DNA replication/repair (Figure 3, Tables 1 and 2).

Identification and Quantitation of Individual Proteins that Cross-Link DNA by Western Blot Analysis

Western blot analysis was used to confirm the identities of the proteins detected by mass spectrometry based proteomics (Tables 1 and 2), and to obtain an estimate of the relative extent to which different proteins become cross-linked to the oligonucleotide duplex in the presence of mechlorethamine. Nuclear protein extracts from CHO and HeLa cells were exposed to 0, 500, or 1000 μ M mechlorethamine, followed by biotin capture of DPCs and SDS-PAGE separation of the proteins as described above. The samples were then transferred to nitrocellulose membranes and probed with commercially available antibodies specific for actin, AGT, nucleolin, histone H4, PARP, Ref-1, GAPDH, Ku, elongation factor 1- α 1, and XRCC1. These proteins were selected based on their identification from the proteomics screen (see above), with the exception of histone H4 and XRCC1 which were not detected by mass spectrometry but chosen because of their propensity to bind chromosomal DNA (29, 30).

First, western blot analysis was used to test for the presence of the actin, AGT, nucleolin, PARP, and XRCC1 proteins in streptavidin-captured DPCs formed by mechlorethamine treatment of nuclear protein extracts from CHO cells. All five proteins were detected, confirming our mass spectrometry results (Figure 4A). By comparing the relative intensity of antigen-specific staining in western blots of affinity captured samples to total nuclear protein extract samples, we were able to estimate the efficiency with which these different proteins became cross-linked to DNA in the presence of mechlorethamine. As the data in Figure 4B reveal, of the five proteins tested, AGT was most efficiently cross-linked to DNA (10% of total protein following exposure to 1000 μ M mechlorethamine, Figure 4B).

A similar analysis was performed using nuclear protein extracts derived from human cervical carcinoma (HeLa) cells. Based on the mass spectrometry results (Table 2) western blots using antibodies specific for actin, AGT, nucleolin, PARP, XRCC, 1Ref-1, Ku, GAPDH, and elongation factor 1- α 1 were employed. As shown in Figure 5A, western blot analysis confirmed protein identities in biotin captured DPCs. Densitometric analysis revealed that the cross-linking efficiencies following exposure to 1000 μ M mechlorethamine was between 4 and 12% of total protein (Figure 5B). Consistent with the mass spectrometry data, western blot analysis failed to detect any histone-DNA cross-links in drug-treated protein extracts (Supplement S-3), in support of our previous finding that purified recombinant histone H4 did not become cross-linked to DNA upon treatment with mechlorethamine (12).

Covalent DPC Formation: MS/MS Analysis of Cys-N7G-EMA

To confirm that mechlorethamine exposure results in the formation of covalent DPCs and to identify the structures of the resulting amino acid-nucleobase conjugates, samples were subjected to neutral thermal hydrolysis and proteolytic digestion to individual amino acids, followed by MS analysis. Based on the previous results for recombinant AGT protein (12), our efforts focused on the detection of *N*-[2-[*S*-cysteinyl]ethyl]-*N*-[2-(guan-7-yl)ethyl] methylamine (Cys-N7G-EMA), the major type of mechlorethamine-induced amino acid-nucleobase conjugate observed in that study, and *N*-[2-[*N*-(lysyl)ethyl]-*N*-[2-(guan-7-yl)ethyl] methylamine (Lys-N7G-EMA). Proteolytic digests were subjected to off-line HPLC separation, and HPLC fractions corresponding to the retention times of Cys-N7G-EMA and Lys-N7G-EMA were collected and analyzed by HPLC-ESI⁺-MS/MS.

Extracted ion chromatograms corresponding to the major MS/MS transition of Cys-N7G-EMA (*m/z* 356 $\rightarrow$ 295) exhibited a prominent peak at $\sim$ 13 min, which matched the HPLC retention time and characteristic SRM transition of synthetic Cys-N7G-EMA (Figure 6). This product was only detected in digests of drug-treated extracts (Figure 6C) and not in a control sample derived from untreated extract (Figure 6B). These results indicate that DNA-protein cross-linking by mechlorethamine takes place between the N7 position of guanine in duplex DNA and side chain sulfhydryls of cysteine residues within proteins, giving rise to *N*-[2-[*S*-cysteinyl]ethyl]-*N*-[2-(guan-7-yl)ethyl]methylamine (Cys-N7G-EMA) cross-links. In contrast, no evidence for the formation of the corresponding lysine conjugate was obtained (results not shown).

Discussion

In the late 1970's, Kohn and colleagues used the biophysical technique of alkaline elution to detect nitrogen mustard-induced DPCs in cells that had been exposed to mechlorethamine and melphalan (10,11). However, no information was available regarding the identities of the proteins involved in DPC formation, nor were the chemical structures of the cross-linked lesions determined. Although previous work in our laboratory showed that mechlorethamine and chlorambucil cross-link the human DNA repair protein AGT to DNA *in vitro* (12), questions remained about the ability of nitrogen mustard drugs to cross-link other nuclear proteins to DNA.

In the present work, mechlorethamine-induced DPCs formed in nuclear protein extracts from mammalian cells were isolated by affinity capture, followed by protein identification by mass spectrometry-based proteomics and western blotting. Analysis of DPCs isolated from mechlorethamine-treated nuclear protein extracts from CHO cells resulted in identification of 15 cross-linked proteins, while analogous experiments conducted using extracts from human cervical carcinoma (HeLa) cells resulted in positive identification of 53 proteins (Tables 1 and 2, respectively). A considerable overlap was observed between the two protein lists. Two thirds of the CHO proteins that were identified following affinity capture of mechlorethamine-

induced DPCs have human homologs that were detected in similar analyses of protein extracts from HeLa cells. The fact that a smaller number of proteins were identified in extracts from hamster cells is not due to more efficient DPC formation with human proteins, but is likely a result of an ascertainment bias resulting from the greater coverage of the human proteome (compared to the hamster proteome) in the databases used for protein identification (~135,000 versus ~1,600 proteins).

Among the proteins participating in DPC formation in the presence of mechlorethamine are those playing a role in cell motility (actin) (31), transcriptional regulation (elongation factor 1- α 1, elongation factor 2, Ref-1, GAPDH), chromatin remodeling (actin, nucleolin) (32-34), DNA supercoiling (topoisomerases I and II) (35), DNA replication (actin) (32), glycolysis, initiation of apoptosis, and vesicle shuttling (33,36-39) (GAPDH), ribosome biogenesis (nucleolin) (40), and DNA repair (nucleolin, AGT, Ku70/86, XRCC1, XRCC5, PARP, Ref-1, Flap endonuclease 1) (13,41-49). Some of these (e.g. actin, tubulin, nucleolin, HSP 90, and heterogeneous nuclear ribonucleoproteins) are high-abundance proteins (50-52) that have a greater probability of cross-linking to DNA in the presence of *bis*-electrophiles due to simple mass action considerations and, once affinity captured, are more likely to be identified by mass spectrometry. However, a subset of proteins identified upon analysis of mechlorethamine-induced DPCs (e.g. AGT, Ref-1, and Flap endonuclease 1) are present within the nucleus at much lower levels (50-52). The latter proteins are involved in DNA replication and repair and probably have a higher propensity to form DPCs because of their strong affinity for DNA and their DNA binding mode that can bring monoalkylated nucleobases in a close proximity to active site residues (e.g. nucleotide flipping for AGT) (53).

The distribution of proteins between functional classes was similar for proteins cross-linked to DNA in the presence of mechlorethamine and the total nuclear proteins found in MS analysis of the human nuclear proteome (54), with the exception of proteins involved in DNA replication and repair which were over-represented among DPC-forming proteins (26% versus 10%). This suggests that proteins of this class are more susceptible to mechlorethamine-mediated DPC formation.

Recently, Qiu and Wang reported the use of formaldehyde cross-linking and tandem mass spectrometry to investigate DNA-protein interactions *in vivo* (55). Following exposure of human acute promyelocytic leukemia cells to formaldehyde, proteins that had become cross-linked to chromosomal DNA were isolated, digested to peptides, and subjected to LC-MS/MS analysis. Identified proteins were then classified according to both cellular localization and function. Of the 780 proteins identified, 305 were classified as nuclear (~39%), and 46 were designated as DNA-binding proteins involved in DNA replication and repair processes (~6 % of all identified proteins) (55). While there are definite similarities in terms of proteins identified as becoming cross-linked to DNA in the presence of both formaldehyde and mechlorethamine (including PARP, Ref-1, and Ku), the observed differences may be the result of varied reactivity/selectivity of the two agents. While formaldehyde is capable of reacting with the amino groups of guanine, adenine, and cytosine in DNA and lysine and arginine in proteins to form DPC lesions (56), alkylation of biomolecules by nitrogen mustards is more selective, reacting primarily with the N7 position of guanine bases and the side chain sulfhydryls of cysteine residues (12). The greater selectivity of nitrogen mustards in forming DPCs could account for the fewer cross-linked proteins identified following mechlorethamine exposure (14 versus 46 with formaldehyde), as well as the identification of a greater percentage of proteins specifically involved in DNA replication and repair (26% versus 6%).

An insight into the chemical structure of mechlorethamine-induced DPCs was provided by HPLC-ESI-MS/MS analyses of total proteolytic digests (Figure 6). These results demonstrate that mechlorethamine-mediated DPCs are covalent in nature, involving the N7 position of

guanine in duplex DNA and the side chain sulfhydryl of cysteine residues within nuclear proteins (*N*-[2-[*S*-cysteinyl]ethyl]-*N*-[2-(guan-7-yl)ethyl]methylamine). The majority of the identified proteins contain at least one cysteine residue (Tables 1 and 2), suggesting that a similar amino specificity may be observed for other protein targets. While we did not observe the formation of the corresponding lysine-guanine cross-links *via* tandem mass spectrometry, cross-linking of other amino acids to DNA by mechlorethamine cannot be ruled out. Studies are currently underway which will allow for direct sequencing of modified peptides to identify the cross-linking sites.

Although the biological implications of DNA-protein cross-linking by mechlorethamine and other *bis*-electrophiles remain to be determined, DPCs are bulky lesions that are expected to lead to genotoxic and cytotoxic effects if left unrepaired. Several types of cellular DPCs have been shown to be relatively long-lived, persisting through several rounds of DNA replication (6,57,58). Thus, DPC formation could result in permanent DNA alterations and other negative consequences, if left unchecked. On the other hand, studies have shown that the majority of DPCs induced by exogenous agents are removed from the genome with time (6), suggesting that these lesions are subject to cellular DNA repair. The mechanism(s) by which this occurs is currently under debate. Based upon the collective experimental evidence (59-61), a model for the repair of covalent DPCs was proposed by Sancar and colleagues (61). In this model, DPC repair is coupled to the replication-dependent proteolysis of cross-linked proteins. Once a DPC is encountered, a component of the stalled replication machinery sends a signal for proteolysis of the covalently attached protein to a peptide, followed by recruitment of proteins involved in nucleotide excision repair (NER). The resulting DNA-peptide cross-link is then removed *via* nucleotide excision repair mechanism. More recently, Nakano *et al.* examined the roles of NER and homologous recombination (HR) in the repair of structurally defined oxidative DPCs *in vitro* and in bacterial cells (62). While NER was involved in the removal of cross-linked proteins of relatively low molecular weights (<14 kDa), proteolytic degradation of DPCs did not contribute to DPC repair as *Escherichia coli* cells deficient in cytosolic ATP-dependent proteases (counterparts of eukaryotic proteasomes) displayed similar results in terms of cell survival. Instead of NER-coupled proteolysis, HR repair was responsible for the removal of DPCs involving oversized proteins (>14 kDa) (62).

Our study demonstrates that numerous nuclear proteins encompassing a variety of cellular functions can become covalently cross-linked to DNA in the presence of the representative nitrogen mustard, mechlorethamine. Although, for practical reasons, we employed much higher nitrogen mustard concentrations than are encountered clinically, it is reasonable to hypothesize that DPCs also form within the cells of individuals treated with nitrogen mustard drugs. Once formed, these DPCs could contribute to the cytotoxicity associated with these agents. For example, unrepaired DPCs could potentially catastrophically interfere with cellular transcription and/or replication, thereby triggering programmed cell death. On the other hand, mis-repair of these lesions could generate lethal mutations or result in chromosomal double-strand breaks, thereby activating programmed cell death. As an initial step towards testing these hypotheses, we are developing new strategies to selectively induce DPCs in a human cell culture model system.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Prof. Anthony E. Pegg (Pennsylvania State University) for generously providing us with hAGT-expressing CHO cells, Prof. James Swenberg (University of North Carolina) for recombinant H4 protein, Brock Matter for help with mass spectrometry experiments, and Bob Carlson (University of Minnesota Cancer Center) for preparing figures

for this manuscript. Funding for this research was from Leukemia Research Foundation, the National Cancer Institute (R01-CA-100670), and a faculty development grant from the University of Minnesota Academic Health Center. R.L. and E.M.R. were partially supported by the NIH Chemistry-Biology Interface Training Grant (T32-GM08700), University of Minnesota Masonic Cancer Center, and University of Minnesota Graduate School.

References

1. Barker S, Weinfeld M, Zheng J, Li L, Murray D. Identification of mammalian proteins cross-linked to DNA by ionizing radiation. *J Biol Chem* 2005;280:33826–33838. [PubMed: 16093242]
2. Zhitkovich A, Voitekun V, Kluz T, Costa M. Utilization of DNA-protein cross-links as a biomarker of chromium exposure. *Environ Health Perspect* 1998;106:969–974. [PubMed: 9703480]
3. Murata-Kamiya N, Kamiya H. Methylglyoxal, an endogenous aldehyde, crosslinks DNA polymerase and the substrate DNA. *Nucleic Acids Res* 2001;29:3433–3438. [PubMed: 11504881]
4. Ewig RA, Kohn KW. DNA damage and repair in mouse leukemia L1210 cells treated with nitrogen mustard, 1,3-bis(2-chloroethyl)-1-nitrosourea, and other nitrosoureas. *Cancer Res* 1977;37:2114–2122. [PubMed: 558823]
5. Kloster M, Kostrhunova H, Zaludova R, Malina J, Kasparkova J, Brabec V, Farrell N. Trifunctional dinuclear platinum complexes as DNA-protein cross-linking agent. *Biochemistry* 2004;43:7776–7786. [PubMed: 15196020]
6. Barker S, Weinfeld M, Murray D. DNA-protein crosslinks: their induction, repair, and biological consequences. *Mutat Res* 2005;589:111–135. [PubMed: 15795165]
7. Rajski SR, Williams RM. DNA Cross-Linking Agents as Antitumor Drugs. *Chem Rev* 1998;98:2723–2796. [PubMed: 11848977]
8. Lawley PD, Brookes P. Molecular mechanism of the cytotoxic action of difunctional alkylating agents and of resistance to this action. *Nature* 1965;206:480–483. [PubMed: 5319105]
9. Osborne MR, Lawley PD. Alkylation of DNA by melphalan with special reference to adenine derivatives and adenine-guanine cross-linking. *Chem Biol Interact* 1993;89:49–60. [PubMed: 8221966]
10. Ross WE, Ewig RA, Kohn KW. Differences between melphalan and nitrogen mustard in the formation and removal of DNA cross-links. *Cancer Res* 1978;38:1502–1506. [PubMed: 565677]
11. Thomas CB, Kohn KW, Bonner WM. Characterization of DNA-protein cross-links formed by treatment of L1210 cells and nuclei with bis(2-chloroethyl)methylamine (nitrogen mustard). *Biochemistry* 1978;17:3954–3958. [PubMed: 568484]
12. Loeber R, Michaelson E, Fang Q, Campbell C, Pegg AE, Tretyakova N. Cross-linking of the DNA repair protein *O*⁶-alkylguanine DNA alkyltransferase to DNA in the presence of antitumor nitrogen mustards. *Chem Res Toxicol* 2008;21:787–795. [PubMed: 18324787]
13. Pegg AE. Repair of *O*⁶-alkylguanine by alkyltransferases. *Mutat Res* 2000;462:83–100. [PubMed: 10767620]
14. Zaidi NH, Liu L, Gerson SL. Quantitative immunohistochemical estimates of *O*⁶-alkylguanine-DNA alkyltransferase expression in normal and malignant human colon. *Clin Cancer Res* 1996;2:577–584. [PubMed: 9816206]
15. Moriwaki S, Nishigori C, Takebe H, Imamura S. *O*⁶-alkylguanine-DNA alkyltransferase activity in human malignant melanoma. *J Dermatol Sci* 1992;4:6–10. [PubMed: 1390458]
16. Lee SM, Rafferty JA, Elder RH, Fan CY, Bromley M, Harris M, Thatcher N, Potter PM, Altermatt HJ, Perinat-Frey T. Immunohistological examination of the inter- and intracellular distribution of *O*⁶-alkylguanine DNA-alkyltransferase in human liver and melanoma. *Br J Cancer* 1992;66:355–360. [PubMed: 1503911]
17. Citron M, Schoenhaus M, Graver M, Hoffman M, Lewis M, Wasserman P, Niederland M, Kahn L, White A, Yarosh D. *O*⁶-methylguanine-DNA methyltransferase in human normal and malignant lung tissues. *Cancer Invest* 1993;11:258–263. [PubMed: 8485648]
18. Bobola MS, Berger MS, Ellenbogen RG, Roberts TS, Geyer JR, Silber JR. *O*⁶-Methylguanine-DNA methyltransferase in pediatric primary brain tumors: relation to patient and tumor characteristics. *Clin Cancer Res* 2001;7:613–619. [PubMed: 11297257]
19. Belanich M, Randall T, Pastor MA, Kibitel JT, Alas LG, Dolan ME, Schold SC Jr, Gander M, Lejeune FJ, Li BF, White AB, Wasserman P, Citron ML, Yarosh DB. Intracellular Localization and

- intercellular heterogeneity of the human DNA repair protein *O*⁶-methylguanine-DNA methyltransferase. *Cancer Chemother Pharmacol* 1996;37:547–555. [PubMed: 8612308]
20. Cai Y, Wu MH, Xu-Welliver M, Pegg AE, Ludeman SM, Dolan ME. Effect of *O*⁶-benzylguanine on alkylating agent-induced toxicity and mutagenicity in Chinese hamster ovary cells expressing wild-type and mutant *O*⁶-alkylguanine-DNA alkyltransferases. *Cancer Res* 2000;60:5464–5469. [PubMed: 11034089]
 21. Jessberger R, Berg P. Repair of deletions and double-strand gaps by homologous recombination in a mammalian *in vitro* system. *Mol Cell Biol* 1991;11:445–457. [PubMed: 1986239]
 22. Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 1976;72:248–257. [PubMed: 942051]
 23. Shin NY, Liu Q, Stamer SL, Liebler DC. Protein targets of reactive electrophiles in human liver microsomes. *Chem Res Toxicol* 2007;20:859–867. [PubMed: 17480101]
 24. Yates JR, Eng JK, McCormack AL, Schieltz D. Method to correlate tandem mass-spectra of modified peptides to amino-acid-sequences in the protein database. *Anal Chem* 1995;67:1426–1436. [PubMed: 7741214]
 25. Eng JK, McCormack AL, Yates JR. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *Journal of the American Society for Mass Spectrometry* 1994;5:976–989.
 26. McDonald WH, Tabb DL, Sadygov RG, MacCoss MJ, Venable J, Graumann J, Johnson JR, Cociorva D, Yates JR III. MS1, MS2, and SQT-three unified, compact, and easily parsed file formats for the storage of shotgun proteomic spectra and identifications. *Rapid Commun Mass Spectrom* 2004;18:2162–2168. [PubMed: 15317041]
 27. Zhang B, Chambers MC, Tabb DL. Proteomic parsimony through bipartite graph analysis improves accuracy and transparency. *J Proteome Res* 2007;6:3549–3557. [PubMed: 17676885]
 28. Balcome S, Park S, Quirk Dorrr DR, Hafner L, Phillips L, Tretyakova N. Adenine-containing DNA-DNA cross-links of antitumor nitrogen mustards. *Chem Res Toxicol* 2004;17:950–962. [PubMed: 15257621]
 29. Horton JK, Watson M, Stefanick DF, Shaughnessy DT, Taylor JA, Wilson SH. XRCC1 and DNA polymerase beta in cellular protection against cytotoxic DNA single-strand breaks. *Cell Res* 2008;18:48–63. [PubMed: 18166976]
 30. Luger K, Richmond TJ. DNA binding within the nucleosome core. *Current Opinion in Structural Biology* 1998;8:33–40. [PubMed: 9519294]
 31. Coulombe RA Jr, Drew GL, Stermitz FR. Pyrrolizidine alkaloids crosslink DNA with actin. *Toxicol Appl Pharmacol* 1999;154:198–202. [PubMed: 9925804]
 32. Zhao K, Wang W, Rando OJ, Xue Y, Swiderek K, Kuo A, Crabtree GR. Rapid and phosphoinositide-dependent binding of the SWI/SNF-like BAF complex to chromatin after T lymphocyte receptor signaling. *Cell* 1998;95:625–636. [PubMed: 9845365]
 33. Mitsuzawa H, Kimura M, Kanda E, Ishihama A. Glyceraldehyde-3-phosphate dehydrogenase and actin associate with RNA polymerase II and interact with its Rpb7 subunit. *FEBS Lett* 2005;579:48–52. [PubMed: 15620689]
 34. Erard MS, Belenguer P, Caizergues-Ferrer M, Pantaloni A, Amalric F. A major nucleolar protein, nucleolin, induces chromatin decondensation by binding to histone H1. *Eur J Biochem* 1988;175:525–530. [PubMed: 3409881]
 35. Champoux JJ. DNA topoisomerases: structure, function, and mechanism. *Annu Rev Biochem* 2001;70:369–413. [PubMed: 11395412]
 36. Tarze A, Deniaud A, Le BM, Maillier E, Molle D, Larochette N, Zamzami N, Jan G, Kroemer G, Brenner C. GAPDH, a novel regulator of the pro-apoptotic mitochondrial membrane permeabilization. *Oncogene* 2007;26:2606–2620. [PubMed: 17072346]
 37. Hara MR, Agrawal N, Kim SF, Cascio MB, Fujimuro M, Ozeki Y, Takahashi M, Cheah JH, Tankou SK, Hester LD, Ferris CD, Hayward SD, Snyder SH, Sawa A. S-nitrosylated GAPDH initiates apoptotic cell death by nuclear translocation following Siah1 binding. *Nat Cell Biol* 2005;7:665–674. [PubMed: 15951807]
 38. Zheng L, Roeder RG, Luo Y. S phase activation of the histone H2B promoter by OCA-S, a coactivator complex that contains GAPDH as a key component. *Cell* 2003;114:255–266. [PubMed: 12887926]

39. Tisdale EJ, Artalejo CR. A GAPDH mutant defective in Src-dependent tyrosine phosphorylation impedes Rab2-mediated events. *Traffic* 2007;8:733–741. [PubMed: 17488287]
40. Ginisty H, Sicard H, Roger B, Bouvet P. Structure and functions of nucleolin. *J Cell Sci* 1999;112(Pt 6):761–772. [PubMed: 10036227]
41. Chen DJ, Park MS, Campbell E, Oshimura M, Liu P, Zhao Y, White BF, Siciliano MJ. Assignment of a human DNA double-strand break repair gene (XRCC5) to chromosome 2. *Genomics* 1992;13:1088–1094. [PubMed: 1505945]
42. Taccioli GE, Gottlieb TM, Blunt T, Priestley A, Demengeot J, Mizuta R, Lehmann AR, Alt FW, Jackson SP, Jeggo PA. Ku80: product of the XRCC5 gene and its role in DNA repair and V(D)J recombination. *Science* 1994;265:1442–1445. [PubMed: 8073286]
43. Tuteja N, Tuteja R, Ochem A, Taneja P, Huang NW, Simoncsits A, Susic S, Rahman K, Marusic L, Chen J. Human DNA helicase II: a novel DNA unwinding enzyme identified as the Ku autoantigen. *EMBO J* 1994;13:4991–5001. [PubMed: 7957065]
44. Nguewa PA, Fuertes MA, Valladares B, Alonso C, Perez JM. Poly(ADP-ribose) polymerases: homology, structural domains and functions. Novel therapeutical applications. *Prog Biophys Mol Biol* 2005;88:143–172. [PubMed: 15561303]
45. Jagtap P, Szabo C. Poly(ADP-ribose) polymerase and the therapeutic effects of its inhibitors. *Nat Rev Drug Discov* 2005;4:421–440. [PubMed: 15864271]
46. Demple B, Harrison L. Repair of oxidative damage to DNA: enzymology and biology. *Annu Rev Biochem* 1994;63:915–948. [PubMed: 7979257]
47. Xanthoudakis S, Curran T. Redox regulation of AP-1: a link between transcription factor signaling and DNA repair. *Adv Exp Med Biol* 1996;387:69–75. [PubMed: 8794196]
48. Levin JD, Demple B. Analysis of class II (hydrolytic) and class I (beta-lyase) apurinic/apyrimidinic endonucleases with a synthetic DNA substrate. *Nucleic Acids Res* 1990;18:5069–5075. [PubMed: 1698278]
49. Marintchev A, Mullen MA, Maciejewski MW, Pan B, Gryk MR, Mullen GP. Solution structure of the single-strand break repair protein XRCC1 N-terminal domain. *Nat Struct Biol* 1999;6:884–893. [PubMed: 10467102]
50. Fountoulakis M, Tsangaris G, Oh J, Maris A, Lubec G. Protein profile of the HeLa cell line. *J Chromatogr A* 2004;1038:247–265. [PubMed: 15233540]
51. Andersen J, Lyon CE, Fox AH, Leung AK, Lam Y, Steen H, Mann M, Lamond AI. Directed proteomic analysis of the human nucleolus. *Curr Biol* 2002;12:1–11. [PubMed: 11790298]
52. Scherl A, Coute Y, Deon C, Calle A, Kindbeiter K, Sanchez J, Greco A, Hochstrasser D, Diaz J. Functional proteomic analysis of the human nucleolus. *Mol Biol Cell* 2002;13:4100–4109. [PubMed: 12429849]
53. Daniels DS, Woo TT, Luu KX, Noll DM, Clarke ND, Pegg AE, Tainer JA. DNA binding and nucleotide flipping by the human DNA repair protein AGT. *Nat Struct Mol Biol* 2004;11:714–720. [PubMed: 15221026]
54. Malmstrom J, Larsen K, Malmstrom L, Tufvesson E, Parker K, Marchese J, Williamson B, Patterson D, Martin S, Juhasz P, Westergren-Thorsson G, Marko-Varga G. Nanocapillary liquid chromatography interfaced to tandem matrix-assisted laser desorption/ionization and electrospray ionization-mass spectrometry: mapping the nuclear proteome of human fibroblasts. *Electrophoresis* 2003;24:3806–3814. [PubMed: 14613209]
55. Qiu H, Wang Y. Exploring DNA-binding proteins with *in vivo* chemical cross-linking and mass spectrometry. *J Proteome Res* 2009;8:1983–1991.
56. Heck HD, Casanova M, Starr TB. Formaldehyde toxicity--new understanding. *Crit Rev Toxicol* 1990;20:397–426. [PubMed: 2198047]
57. Cupo DY, Wetterhahn KE. Binding of chromium to chromatin and DNA from liver and kidney of rats treated with sodium dichromate and chromium(III) chloride *in vivo*. *Cancer Res* 1985;45:1146–1151. [PubMed: 2578874]
58. Oleinick NL, Chiu SM, Ramakrishnan N, Xue LY. The formation, identification, and significance of DNA-protein cross-links in mammalian cells. *Br J Cancer* 1987;135–140. [PubMed: 3545275]

59. Quievryn G, Zhitkovich A. Loss of DNA-protein crosslinks from formaldehyde-exposed cells occurs through spontaneous hydrolysis and an active repair process linked to proteasome function. *Carcinogenesis* 2000;21:1573–1580. [PubMed: 10910961]
60. Baker DJ, Wuenschell G, Xia L, Termini J, Bates SE, Riggs AD, O'Connor TR. Nucleotide excision repair eliminates unique DNA-protein cross-links from mammalian cells. *J Biol Chem* 2007;282:22592–22604. [PubMed: 17507378]
61. Reardon JT, Cheng Y, Sancar A. Repair of DNA-protein cross-links in mammalian cells. *Cell Cycle* 2006;5:1366–1370. [PubMed: 16775425]
62. Nakano T, Morishita S, Katafuchi A, Matsubara M, Horikawa Y, Terato H, Salem AM, Izumi S, Pack SP, Makino K, Ide H. Nucleotide excision repair and homologous recombination systems commit differentially to the repair of DNA-protein crosslinks. *Mol Cell* 2007;28:147–158. [PubMed: 17936711]

Figure 1.

Concentration-dependent formation of DPCs in nuclear protein extracts prepared from CHO (A) and HeLa (B) cells following exposure to mechlorethamine. Nuclear protein extracts (500 μg) and 5'-biotinylated double-stranded oligodeoxynucleotides (3.12 nmol) were incubated with 0-1000 μM mechlorethamine (lanes 3-8), subjected to biotin capture enrichment, and resolved by 10% SDS-PAGE. The gels were stained with SimplyBlue SafeStain to visualize the cross-linked proteins.

Figure 2.

HPLC-ESI⁺-MS/MS identification of tryptic peptides resulting from biotin capture of mechlorethamine-induced DNA-protein cross-links involving GAPDH (**A**), PARP (**B**), and nucleolin (**C**).

Figure 3.
Cellular functions of CHO and human proteins that form DPCs in the presence of mechlorethamine.

Figure 4.

Western blot analysis of mechlorethamine-induced DNA-protein cross-links in nuclear extracts from CHO cells. **(A)** Nuclear extract proteins from CHO cells were incubated with 0 (lane 1), 500 μ M (lane 2), or 1000 μ M mechlorethamine (lane 3) in the presence of 5'-biotinylated double-stranded oligodeoxynucleotides. Following biotin capture enrichment, cross-linked proteins were resolved by SDS-PAGE and transferred to nitrocellulose membranes. Western blots were performed using primary antibodies specific for: actin, *O*⁶-alkylguanine DNA alkyltransferase (AGT), nucleolin, poly (ADP-ribose) polymerase (PARP), and X-ray repair cross-complementary protein 1 (XRCC1). Untreated nuclear extract (~25 μ g) was loaded as a positive control (lane 4). In cases where the samples were not loaded in adjacent gel lanes, protein signals are outlined separately. **(B)** Densitometric analysis of western blots to estimate the extent of protein cross-linking to DNA in the presence of mechlorethamine.

Figure 5. Western blot analysis of mechlorethamine-induced DNA-protein cross-links in nuclear extracts for HeLa cervical carcinoma cells: (A) Nuclear extract proteins from HeLa cells were incubated with 5'-biotinylated double-stranded oligodeoxynucleotides in the presence of 0 (lane 1), 500 μ M (lane 2), or 1000 μ M mechlorethamine. Following biotin capture enrichment, proteins participating in DPC formation were resolved by 10 or 12% SDS-PAGE and transferred to nitrocellulose membranes. Western blots were performed using primary antibodies specific for actin, nucleolin, poly (ADP-ribose) polymerase (PARP), DNA-(apurinic or apyrimidinic site) lyase (Ref-1), ATP-dependent DNA helicase subunit 2 (Ku), glyceraldehyde-3-phosphate dehydrogenase (GAPDH), elongation factor 1- α 1 (EF-1 α 1), *O*⁶-alkylguanine DNA alkyltransferase (AGT), and X-ray repair cross-complementary protein 1 (XRCC1). Untreated nuclear extract (~25 μ g) was loaded as a positive control (lane 4). In cases where the samples were not loaded in adjacent gel lanes, protein signals are outlined separately. (B) Densitometric analysis of western blots to estimate the extent of protein cross-linking to DNA in the presence of mechlorethamine.

Figure 6.

HPLC-ESI⁺-MS/MS analysis of *N*-[2-*[S*-cysteiny]ethyl]-*N*-[2-(guan-7-yl)ethyl]methanamine (Cys-N7G-EMA) in proteolytic digests. Nuclear protein extracts from HeLa cells were incubated with mechlorethamine in the presence of double-stranded DNA, and the resulting DPCs were subjected to thermal and enzymatic hydrolysis to release nucleic acid-nucleobase conjugates (synthetic Cys-N7G-EMA (A), enzymatic digests of HeLa nuclear extracts following incubation in the presence of DNA (negative control) (B), enzymatic digests of HeLa nuclear extracts following incubation with 1000 μ M mechlorethamine in the presence of DNA (C).

Scheme 1.

Nitrogen mustard-mediated formation of DNA-protein cross-links.

Scheme 2.

Experimental scheme for biotin capture methodology used to isolate DNA-protein cross-links from nuclear protein extracts incubated with mechlorethamine in the presence of double-stranded DNA.

Table 1
CHO nuclear extract proteins that form DPCs in the presence of 500 μ M mechlorethamine

Swiss-Prot ID	Protein	Peptide Sequences	Spectra	Classification	Number of Cysteines
P48975	Actin	19	276	Architectural/Structural	6
P69893	Beta tubulin	24	144		8
P68361	Alpha tubulin	15	125		12
P68362	Alpha tubulin II	14	103		12
Q9J155	Plectin-1	23	34		5
Q60399	Nucleolin (C23)	22	255	Chromatin Regulators	1
P17244	Glyceraldehyde-3-phosphate dehydrogenase	9	109	Cellular Homeostasis/Cell Cycle	4
P46633	Heat shock protein HSP 90-alpha	24	154		7
P19378	Isoform 1 of Heat shock cognate 71 kDa protein	19	133		4
P62629	Elongation factor 1-alpha 1	23	196	Transcription Regulators/RNA Splicing	6
P09445	Elongation factor 2	23	98		17
Q07050	DNA topoisomerase I	14	69	DNA Replication/Repair	8
P16455	Methylated-DNA--protein-cysteine methyltransferase (AGT)	4	41		5
P41515	DNA topoisomerase 2-alpha	18	34		13
Q9R152	Poly [ADP-ribose] polymerase 1 (PARP)	11	18		16

Table 2
HeLa nuclear extract proteins that form DPCs in the presence of 500 μ M mechlorethamine

Swiss-Prot ID	Protein	Peptide Sequences	Spectra	Classification	Number of Cysteines
P60709	Actin	12	109	Architectural/Structural	6
P02545	Isoform A of Lamin-A/C	25	98		5
Q9BQE3	Tubulin alpha-1C chain	9	49		12
Q14978	Isoform Beta of Nucleolar phosphoprotein p130	7	39		0
P19338	Nucleolin (C23)	18	197	Chromatin Regulators	1
Q9Y5B9	FACT complex subunit SPT16	35	127		10
Q08945	FACT complex subunit SSRP1	20	116		6
Q01105	Isoform 1 of Protein SET	7	54		0
Q13283	Ras GTPase-activating protein-binding protein 1	8	43		1
Q9UQE7	Structural maintenance of chromosomes protein 3	22	43		7
P04406	Glyceraldehyde-3-phosphate dehydrogenase	9	109	Cellular Homeostasis/Cell Cycle	3
Q09666	AHNAK nucleoprotein isoform 1	48	97		20
Q9P258	Protein RCC2	14	74		13
Q92688	Isoform 1 of Acidic leucine-rich nuclear phosphoprotein 32 family member B	7	60		3
Q99459	Cell division cycle 5-like protein	16	52		4
P11142	Isoform 1 of Heat shock cognate 71 kDa protein	12	47		4
P07900	Heat shock protein HSP 90-alpha	12	40		7
Q16666	Isoform 1 of Gamma-interferon-inducible protein Ifi-16	11	35		8
P49736	DNA replication licensing factor MCM2	8	17		12
P61978	Isoform 2 of Heterogeneous nuclear ribonucleoprotein K	22	254	Transcription Regulators/RNA Splicing	5
P06748	Isoform 2 of Nucleophosmin	11	217		3
Q15233	Non-POU domain-containing octamer-binding protein	27	189		2
P52272	Isoform 1 of Heterogeneous nuclear ribonucleoprotein M	36	167		5
Q08211	ATP-dependent RNA helicase A	44	148		23
P23246	Isoform Long of Splicing factor, proline- and glutamine-rich	20	132		2
P26599	Isoform 1 of Polypyrimidine tract-binding protein 1	20	132		3

Swiss-Prot ID	Protein	Peptide Sequences	Spectra	Classification	Number of Cysteines
Q92841	DEAD box polypeptide 17 isoform 1	27	116		12
P09651	Isoform A1-B of Heterogeneous nuclear ribonucleoprotein A1	14	115		2
Q12906	Interleukin enhancer-binding factor 3	26	89		7
P60842	Eukaryotic initiation factor 4A-1	22	82		4
O43390	Heterogeneous nuclear ribonucleoprotein R	17	78		5
O60506	Isoform 3 of Heterogeneous nuclear ribonucleoprotein Q	20	77		4
P22626	Isoform B1 of Heterogeneous nuclear ribonucleoproteins A2/B1	12	75		1
P68104	Elongation factor 1-alpha 1	9	71		6
P31943	Heterogeneous nuclear ribonucleoprotein H	9	57		5
Q15459	Splicing factor 3 subunit 1	13	55		1
P17844	Probable ATP-dependent RNA helicase DDX5	18	54		9
Q16630	Isoform 1 of Cleavage and polyadenylation specificity factor subunit 6	10	39		3
P43243	Matrin-3	7	20		9
P12956	ATP-dependent DNA helicase 2 subunit 1 (Ku70)	26	186	DNA Replication/Repair	5
P27695	DNA-(apurinic or apyrimidinic site) lyase (Ref-1)	14	128		7
P13010	ATP-dependent DNA helicase 2 subunit 2 (Ku86)	24	117		10
P11388	DNA topoisomerase 2-alpha	39	91		14
P11387	DNA topoisomerase 1	22	80		8
Q02880	DNA topoisomerase 2-beta	26	70		17
P09874	Poly [ADP-ribose] polymerase 1 (PARP)	28	69		14
P39748	Flap endonuclease 1	6	24		6
Q16531	DNA damage-binding protein 1	15	23		22
Q9Y230	RuvB-like 2	8	20		3
P43246	DNA mismatch repair protein Msh2	7	19		13
P49959	Isoform 1 of Double-strand break repair protein MRE11A	7	18		6
P16455	Methylated-DNA--protein-cysteine methyltransferase (AGT)	2	5		5
O60216	Double-strand-break repair protein rad21 homolog	2	5		5