The Extreme C-terminus of Xeroderma pigmentosum group G protein Mediates Protein-Protein Interactions

Kiran Rangaraj
Office of Science, Science Undergraduate Laboratory Internship Program
Correspondence: kiranrangaraj@yahoo.com

University of California, Berkeley

Lawrence Berkeley National Laboratory
Berkeley, California

August 24, 2009

Prepared in partial fulfillment of the requirement of the Office of Science, Department of Energy's Science Undergraduate Laboratory Internship under the direction of Kelly S. Trego and Priscilla K. Cooper in the Life Sciences division at Lawrence Berkeley National Laboratory.

TABLE OF CONTENTS

ABSTRACT

The Extreme C-terminus of Xeroderma pigmentosum group G protein Mediates Protein-Protein Interactions.

Kiran Rangaraj (University of California, Berkeley, Berkeley, CA 94720), Kelly S. Trego and Priscilla K. Cooper (Lawrence Berkeley National Laboratory, Berkeley, CA 94720).

Organisms across the evolutionary scale are equipped with complex and interconnected DNA repair pathways that are regulated by multifunctional proteins. These proteins mediate interactions by conformational changes and protein hand-offs in order to coordinate lesion detection and removal with vital cellular processes such as DNA replication, transcription and recombination. Mutations that disrupt repair protein functioning can lead to genomic instability, developmental and immunological abnormalities, and cancer and aging. Xeroderma pigmentosum group G (XPG) is one such multifunctional protein that plays a critical role in maintaining human genome stability. Point mutations in the XPG gene gives rise to an inherited photosensitive disorder, Xeroderma pigmentosum (XP) and truncation mutations cause the profound neurological and developmental disorder Cockayne syndrome (CS) combined with XP. The molecular basis of XPG in XP is well understood because XPG contains structure specific 3' endonuclease activity that is critical to the repair of ultraviolet-damaged DNA in the nucleotide excision repair (NER) pathway. However, the clinical features of CS in XPG-CS patients are difficult to explain on the basis of defects in NER, which suggests that XPG possesses several poorly understood roles that are regulated by its unstructured non-enzymatic recognition (R) and carboxyl (C) terminal domains. These domains have been shown to mediate interactions with over fifteen proteins from multiple repair pathways. Studies conducted on these regions have identified novel scaffolding roles for XPG in transcription-coupled (TCR) and base excision (BER) repair, and recently a replication-associated function with proteins that process damaged replication forks. How XPG is involved in multiple pathways is of considerable interest. Considering the role of the XPG C-terminus in protein-protein interaction, this study involved bacterially expressing and purifying three sequential C-terminal subdomain constructs and screening interactions with seven proteins representing roles in different DNA replication and repair pathways. All seven proteins were found to interact with the same region of the C-terminus, which provides information critical towards identifying the amino acids uniquely required by each protein partner. This will allow one to genetically dissect the molecular basis of XPG in order to elucidate and remedy its involvement in the complex disease phenotype of XP-CS.

1. INTRODUCTION

Genomic integrity is persistently threatened by endogenous (internal) damage from cellular metabolism and environmental sources, including radiation and chemicals. This damage causes an extensive variety of altered DNA bases, as well as single-strand (ssDNA) and double-strand (dsDNA) DNA breaks. To counter this damage, organisms across the evolutionary scale are equipped with complex and interconnected pathways that are regulated by multifunctional proteins and the transient protein complexes or assemblies they often form by conformational changes and protein hand-offs. These proteins coordinate lesion detection and removal with vital processes such as DNA replication, transcription and recombination. Failure in coordination can lead to permanent genetic changes and tumorigenesis, or to cell death, senescence and premature aging, both driven by endogenous damage and exacerbated by environmental mutagens. Elucidating the precise mechanism by which these processes are carried out will provide a considerable increase in our understanding of how DNA repair processes maintain the integrity of the human genome.

A large variety of potentially mutagenic DNA lesions can be produced upon exposure to carcinogenic compounds and ultraviolet (UV) radiation. This includes bulky base adducts that are produced from carcinogens such as benzo[a]pyrene, which is present in automobile exhaust fumes and tobacco smoke. Also included are photoproducts produced the UV-promoted covalent cross-linking of adjacent pyrimidine bases, which distort the DNA-helix. These different kinds of lesions are recognized and repaired by the versatile nucleotide excision repair (NER) pathway that is conserved in eukaryotes (Figure 1) [1]. NER consists of an initial recognition step, followed by removal of a short single-strand of DNA that contains the lesion. This results in a single-strand gap that is subsequently repaired by a DNA polymerase, which uses the undamaged strand as a template. The initial recognition step can occur in two significantly different forms, providing two sub-pathways to NER. These are transcription-coupled repair (TCR) and global genome repair (GG-NER) [3]. TCR functions to repair the actively transcribed strand of genes and at a rate of about four times that of GG-NER, which repairs both the transcribed and nontranscribed strands [4]. Rapid repair of damage in the

transcribed strand is imperative, as an unrepaired lesion leads to both reduced transcription of critical genes and provides a potent signal for apoptosis [5].

Much of our understanding of NER has come from the study of cells from individuals with hereditary mutations in GG-NER or TCR genes that cause clinical phenotypes [6]. This includes the sunlight sensitive and skin cancer prone disorder xeroderma pigmentosum (XP), which can be caused by a mutation in one of seven key GG-NER genes, XPA through to XPG, plus one further XP-related gene, XPV [7]. Mutations in critical TCR genes can result in Cockayne syndrome (CS) (Figure 2). The clinical phenotype of CS is sun sensitivity, severe physical and mental retardation and pronounced cachexia (wasting) and these symptoms result in childhood mortality in the more severe cases [9]. A majority of CS cases are caused by mutations in either the CSA or CSB genes, which have specialized functions in the early recognition step of TCR. The remaining cases are caused by mutations in the GG-NER genes XPB, XPD and XPG that also function in the later steps of TCR. Mutations in these genes result in CS combined with XP, though how these mutations give rise to this clinical phenotype is poorly understood. Recent studies have suggested that the CS phenotype is related to the high levels of oxidative stress occurring in the central nervous system, which implies that CS proteins may be additionally involved in the recognition and repair of oxidatively damaged DNA (base excision repair, BER) [9]- [11]. BER is a highly coordinated eukaryotic repair pathway that eliminates oxidized and alkylated bases and apurinic/apyrimidic sites from DNA.

Implications for BER-deficiency in CS are evident in the XPG (ERCC5) gene, which encodes for an 1186 residue containing structure-specific endonuclease [12], [13]. Recent studies have demonstrated XPG protein to have multiple poorly understood roles including transcription and the repair of oxidative base damage, in addition to its well documented involvement in NER [12]-[14]. XPG protein has been observed to correct cells from individuals showing deficiency of XP complementation group G, which are highly UV sensitive and NER-deficient [15], [16]. XPG functions in the late stage of NER by incising the damaged DNA strand in a transcription bubble structure zero to two nucleotides from the ssDNA-dsDNA junction, and is a member of the Flap EndoNuclease-1 (FEN-1) family of structure-specific nucleases (Figure 3) [15]-[17].

The structure of XPG protein is undetermined, but structures of family members are observed to form an α/β 'saddle' structure built from two regions, N and I, which together form the nuclease active site and which are separated by a flexible loop of approximately 70 amino acids [18]-[21]. The XPG sequence differs from the family by containing a large spacer region of over 600 amino acids (the Recognition, R-domain) and a C-terminus that is extended by approximately 145 residues. The large R- and C-terminal domain sequences do not contain clear sequence similarity to any known structural motifs. While point mutations that perturb the nuclease function of the protein cause an XP phenotype, mutations that truncate the protein cause XP-CS [22]-[24]. A seven-residue deletion in the R-domain and a truncation of 260 residues in the C-terminus can give rise to the XP-CS phenotype [22], [24], [25]. This indicates that both nonenzymatic regions have critical functions in the cell, which could provide valuable insight into the molecular mechanisms behind TCR pathway progression and the CS phenotype if elucidated, and this is the primary research focus of the Cooper Laboratory of the Life Sciences Division at Lawrence Berkeley National Laboratory.

Studies conducted on these nonenzymatic regions have identified novel scaffolding roles for XPG in TCR and BER, and most recently a replication-associated function involving interactions with proteins that process damaged replication forks (Cooper Lab, unpublished results). With regard to TCR, the R-domain is required for XPG to bind to NER bubble substrates [26]-[28]. The C-terminus extension of the bound XPG then mediates an interaction with CSB, which stimulates both bubble DNA binding and ATPase activity of CSB [28]. This finding implicates XPG function in the early recognition stage of TCR that requires partnership with CSB protein. This early stage function is also suggested by interaction of XPG with elongating RNA polymerase II (RNAP II) *in vivo*, and with stalled RNAP II *in vitro* [28]. Notably, the C-terminal region of XPG has also been observed to bind to the transcription and repair factor TFIIH, which is a multi-protein complex required for transactivation of nuclear receptors in NER (Figure 4) [29]. Mutants of XPG protein lacking the C-terminus due to XP-CS mutations did not form a stable complex with TFIIH, which caused a dissociation of the cdk-activating kinase (CAK) subunit from the TFIIH core, while XPG mutant proteins with the classical XP phenotype bind efficiently [29].

Cooper Lab studies have identified several implications for XPG involvement in BER. Notably, XPG-CS cells (lacking full-length XPG) are significantly hyper-sensitive to oxidizing agents such as peroxide and ionizing radiation, whereas point mutations in XPG that inactivate nuclease activity retain normal levels of resistance to oxidizing agents (Cooper Lab, unpublished results). With regard to the early steps of BER, the R- and C-terminal regions of XPG have been found to interact with and stimulate the incision activity of multiple BER enzymes. XPG has been found to synergistically stimulate the enzymatic activities of the DNA glycosylase proteins nth endonuclease III-like 1 (NTH1) and apurinic/ apyrimidic endonuclease-1 (APE1), by loading them onto their substrates, promoting turnover, and facilitating handoff interactions between them (Cooper Lab, unpublished results). The C-terminus alone has also demonstrated the ability to similarly stimulate the activities of the Nei endonuclease VIII-like 1 (NEIL1) and Nei like 2 (NEIL2) glycosylases, which are implicated in replication-associated and transcription-coupled BER respectively (Cooper Lab, unpublished results) [30]. Consistent with this biochemical data, whole cell extracts (WCEs) from three different XPG-CS patient cell lines were shown severely defective in incision at a dihydrouracil lesion as compared to wild-type cells. The addition of XPG to the XPG-CS WCEs resulted in near complete restoration of incision activity (Cooper Lab, unpublished results). These novel roles for the R-domain and C-terminus of XPG suggest that non-catalytic regions of CS proteins are required to play a scaffolding role to mediate protein hand-offs required for effective DNA repair.

Cooper Lab studies have additionally suggested a novel replication-associated role for XPG by demonstrating stable C-terminal regulated interactions with several proteins involved in double-strand break (DSB) repair and signaling or processing damaged replication forks, including both Werner (WRN) and Bloom (BLM) syndrome helicases (Cooper Lab, unpublished results). These interactions have been demonstrated using both a directed search by Far Western and an unbiased mass spectrometry characterization of proteins that co-immunoprecipitate with XPG from human cells.

Overall, this indicates that the XP-CS phenotype might arise from perturbing interactions between the R- and C-terminal domains of XPG and any one of its numerous protein partners from TCR, BER and DSBR (Figure 5). This evidence also supports the concept that one possible manner in which XPG proteins multiple functions are precisely regulated is through protein-protein interaction. Genetic dissection of the role of XPG in these pathways would strengthen the notion that the protein interaction partners of XPG regulate its activity. However, whether these protein partners utilize the same interaction surfaces remains to be shown. In addition, precise genetic dissection of the multiple roles of XPG would require information regarding a single or multiple amino acid residue mutation that affects only one of the known XPG interactions while leaving others relatively undisturbed. Considering the role of the XPG C-terminus in protein-protein interaction thus far, this study involved screening the C-terminus with three bacterially expressed and purified subdomain constructs of this region with the rationale that this would yield information regarding the importance of certain specific regions in mediating interactions with various proteins (Figure 6). The recombinant C-terminal subdomains were screened with seven proteins, WRN, BLM, replication protein A (RPA), NEIL1, NEIL2, NTH1, and APE1 that are representative of various DNA metabolic pathways. Results provide information critical towards identifying the amino acids uniquely required by each protein interaction partner, which will allow one to genetically dissect the molecular basis of XPG in the complex disease phenotype of XP-CS.

2. MATERIALS AND METHODS

i. Cloning of MBP-XPG C-terminal Subdomains

The DNA sequence coding for each C-terminal construct (E1: 1012-1055, E2: 1055-1107, E3: 1107-1186) was amplified by polymerase chain reaction (PCR) from a template vector pGAZ-MBP Exon-15, containing the entire XPG C-terminal DNA sequence (1007-1186), using restriction enzymes EcoRI and HindIII (New England BioLabs (NEB)). Each C-terminal DNA fragment was then subcloned into a Kanamycin resistant pGAZ-MBP vector, which appended the fragments with an N-terminal six-histidine tag and maltose binding protein (MBP) fusion sequence. Ligation of the XPG C-terminal DNA fragments with the pGAZ-MBP cloning vector involved digesting the vector at its multiple cloning site (MCS) using the same two restriction

enzymes, EcoRI and HindIII, in order to generate the palindrome complementary sequence required for ligation. The digested cloning vector was treated with the enzyme Antarctic Phosphatase (NEB), which catalyzed the removal of the 5' phosphate group from DNA in order to prevent self-ligation/ re-circularization. The amplified XPG C-terminal DNA fragments and the digested pGAZ-MBP vector were then gel purified using 1% Agarose TAE (Tris-acetate-ethylenediaminetetraacetic acid (EDTA)) containing 0.02 µg/mL ethidium bromide, and further purified by extraction from agarose and all other impurities using the QIAquick Gel Extraction Kit (Qiagen). Concentrations of the isolated C-terminal fragments and pGAZ-MBP vector were quantified from comparison to known quantities of a High and Low DNA Mass Ladder (Invitrogen) on 1% Agarose TAE. Ligation was then performed by adding Quick T4 DNA Ligase (NEB) to a ten to one molar ratio of the purified restriction digested XPG C-terminal DNA fragments and the pGAZ-MBP cloning vector. The ligated plasmid vectors were then transformed into competent DH5α subcloning efficient E. coli cells (Invitrogen) by heat shocking for one minute at 37˚C. The transformed bacteria were incubated for one hour in S.O.C. growth medium (Invitrogen) at 37˚C while shaking at 250 revolutions per minute (rpm) to obtain maximal transformation efficiency and spread onto Kanamycin (0.03 mg/mL) selective Lysogeny broth (LB) plates, which selected for the growth of only Kanamycin resistant (successfully transformed) bacterial colonies when incubated overnight at 37˚C. Bacterial colonies on each XPG C-terminal plate were selected for and separately underwent plasmid purification via the QIAprep Spin Miniprep Kit (Qiagen). Transformation efficiency and purity of plasmid purification were both assessed by restriction digesting the different pools of purified plasmid using EcoRI and HindIII, and resolving by 1% Agarose TAE containing 0.02 µg/mL ethidium bromide in comparison to uncut, circular parental and recombinant pGAZ-MBP vectors. The DNA was visualized in the gel by the addition of ethidium bromide, which binds strongly to DNA by intercalating between the bases. The fluorescent properties of ethidium bromide allow it to absorb UV light and visibly transmit this energy.

ii. Recombinant Protein Expression of MBP-XPG C-terminal Subdomains

Optimal pools of purified pGAZ-MBP vector containing the XPG C-terminal subdomains of interest (E1, E2, & E3) were transformed into E. coli Rosetta 2 competent cells (Novagen)

possessing Chloramphenicol resistance. Transformation was conducted by incubating the plasmids and Rosetta 2 cells for 5 minutes over ice, heat shocking for one minute at 37˚C, and incubating for one hour in S.O.C. growth medium at 37˚C while shaking at 250 rpm. The cells were then spread onto LB plates containing working concentrations of Kanamycin (0.03 mg/mL) and Chloramphenicol (0.034 mg/mL), which selected for the growth of bacterial colonies successfully transformed with the recombinant pGAZ-MBP vector. Two six milliliter bacterial cultures were grown per MBP-XPG C-terminal construct by inoculating LB media containing working concentrations of Kanamycin and Chloramphenicol with 2-3 bacterial colonies, and incubating overnight at 37˚C while shaking at 250 rpm. A larger one to one hundred dilution of the overnight culture in one liter of LB media containing working concentrations of Kanamycin and Chloramphenicol was made, and allowed to grow at 37˚C while shaking at 250 rpm until an optical density reading of 0.5 at 600 nm (logarithmic phase) was reached. Protein expression of MBP-XPG E1 was induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.1 mM at 37˚C while shaking at 250 rpm for 4 hours. Protein expression of MBP-XPG E2 and MBP-XPG E3 was induced with 0.1 mM IPTG for 16 hours at 16˚C while shaking at 250 rpm. Cells were harvested by ultracentrifugation at 6000 rpm for 20 minutes at 4˚C in two balanced 500 milliliter aliquots from each one liter bacterial cell culture.

Cell pellets were resuspended in 20 milliliters of buffer A (NaH_2PO_4 (pH 8.0) containing 300 mM NaCl, 10 mM Imidazole, 0.1% nonyl phenoxylpolyethoxylethanol (NP-40), 0.5 mM phenylmethanesu1fonylfluoride (PMSF), 1 µg/mL aprotinin, 1 µg/mL leupeptin, and 0.5 µg/mL pepstatin), incubated on ice for 30 minutes and lysed under 20,000 pound force per square inch (psi) of pressure using Constant Cell Disruption Systems high-shear mechanical cell disruptor. Ten milliliters of buffer A were used to rinse the constant cell disruptor and combined with the lysed samples, which were then clarified by ultracentrifugation at 12,500 rpm for 25 minutes. The soluble extract was retained for purification and protein expression was confirmed by resolving the proteins upon 12% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) in Tris-glycine SDS running buffer (deionized water with 25 mM Tris-base, 0.19 M glycine (Bio-Rad), and 3.4 mM of SDS). The resolved proteins were visualized by staining with Bio-Safe Coomassie Stain (Bio-Rad).

iii. Affinity Purification of MBP-XPG C-terminal Subdomains

HisPur Cobalt Resin (Thermo Scientific) was equilibrated with buffer A in a one to one volumetric ratio. Eight milliliters of the buffer A and bead resin mixture was added to the soluble protein extract of MBP-XPG E1. Nickel-nitrilotriacetic acid (Ni-NTA) Superflow Resin (Qiagen) was equilibrated with buffer A in a one to one volumetric ratio, and ten milliliters of the 50% bead slurry was added to each of the soluble protein extracts of MBP-XPG E2 and MBP-XPG E3. The proteins were allowed to bind to the resin for one hour at 4˚C with gentle agitation via affinity between the proteins polyhistidine-tag and the cobalt or nickel ion resin. The bead slurries were loaded onto separate columns (15 mm x 150 mm, Bio-Rad) that were clamped upright and previously rinsed with buffer A, and allowed to gravity settle over 30 minutes to pack into a 4 or 5 milliliter resin-bed while collecting the flow-through. Each column was loaded onto a low pressure liquid chromatography system (Bio-Rad BioLogic LP) and washed with 10 column volumes of buffer A containing 20 mM imidazole. The bound protein was eluted with 10 column volumes of buffer A containing 0.5 M imidazole, monitored by UV absorbance at 280 nm, and collected in 20 half-column volume fractions. All steps of the protein purification were performed at 4˚C. The peak fractions were identified by Bio-Safe Coomassie stained SDS-PAGE, pooled, concentrated in storage buffer (20 mM Tris-HCl (pH 8.0), 500 mM NaCl, 1 mM ethylenediaminetetraacetic acid (EDTA), 5% glycerol, 0.5 mM PMSF, 1 µg/mL aprotinin, 1 µg/mL leupeptin, and 0.5 µg/mL pepstatin) using Amicon Ultra-4 centrifugal filter devices (Millipore) and stored at −80˚C. Protein concentrations were quantified from comparison with BSA standards on Bio-Safe Coomassie stained 4-12% gradient SDS-PAGE.

iv. Additional Proteins

Full length XPG protein was purified in the Cooper lab as previously described, and MBP-XPG Exon 15 and MBP were obtained from Miaw-Sheue Tsai of the Expression and Molecular Biology (EMB) Core (Life Sciences Division, LBNL, Berkeley, CA) [28]. All proteins used for interaction mapping with XPG were obtained from different parties of the Cooper Lab: Altaf H. Sarker provided the four glycosylase proteins NEIL1, NEIL2, NTH1, and APE1; Susan E. Tsutakawa provided RPA; Kelly S. Trego provided BLM and WRN proteins. WRN protein was purified as previously described [31].

v. Protein-Protein Interaction Assay

Far Western analysis was performed by denaturing 3-4.5 µg of each polypeptide (full-length XPG, XPG domain constructs, MBP and any additional control proteins) by boiling for five minutes in a 1:6 volumetric dilution of 6X SDS-sample buffer (70% 0.5 M Tris, pH 6.8, and 30% glycerol, with 0.35 M SDS, 0.6M (D,L)-1,4-dithiothreitol (DTT) from Soltec Ventures, Inc., and 0.18 mM Bromophenol Blue dye (BPB)) upon a 95˚ C heat block. Proteins were then resolved by SDS-PAGE using a precast 1.0 mm x 10 well, Novex® 4-12% Tris-Glycine Gel (Invitrogen). Molecular weight quantification of the proteins was achieved using Kaleidoscope Prestained Standards (Bio-Rad). Proteins were transferred overnight to Trans-Blot nitrocellulose filters (Bio-Rad) using the corresponding Mini Trans-Blot Electrophoretic Transfer Cell (Bio-Rad) in standard Tris-Glycine transfer buffer (deionized water containing 20% methanol, with 25 mM Tris-base, and 190 mM glycine) at 4˚ C. The filters were briefly immersed in 1X Phosphate Buffered Saline (PBS) (Omni Pur) then Ponceau S Solution (Sigma-Aldrich) to provide temporary protein visualization and gauge the efficiency of transfer. Brief subsequent immersions in PBS allowed destaining of the background for increased visual clarity. Filters were then immersed twice in denaturation buffer (6 M guanidine-HCl in PBS) for 5 minutes and incubated six times for 10 minutes each in serial dilutions (1:1) of denaturation buffer in PBS to allow progressive protein renaturation. Filters were blocked in PBS containing five percent nonfat powdered milk for three hours at room temperature to prevent non-specific background binding of the antibodies to the membrane, prior to being incubated in full-length WRN, BLM, RPA, NEIL1, NEIL2, NTH1, or APE1 protein probes (10.0 pmole / mL) in PBS supplemented with 0.5% powdered milk, 0.1 or 0.5% Tween-20 (Sigma-Aldrich), 1.0 mM DTT, and 0.5 mM PMSF overnight at 4˚ C with gentle agitation. Following overnight incubation, the filters were extensively washed four times for 10 minutes each in PBS containing 0.1-1.0% Tween-20, after which conventional Western blotting was performed to detect the presence of the protein probes using mouse monoclonal αWRN (4H12: Abcam) at a 1:666 dilution, goat polyclonal αBLM (K-20: Santa Cruz Biotechnology) at a 1:200 dilution, mouse monoclonal αRPA (RPA70-9: Calbiochem) at a 1:40 dilution, rabbit polyclonal αNEIL1 (ab21337: Abcam) at a 1:1000 dilution, mouse monoclonal αNEIL2 (ab56577: Abcam) at a 1:600 dilution, rabbit polyclonal αNTH1 (ab93: Abcam) at a 1:3000 dilution, or rabbit polyclonal $\boldsymbol{\alpha}$APE1 (sc334: Santa Cruz

Biotechnology) at a 1:1000 dilution, as primary antibody. After one-hour incubation with the primary antibody, the filters were washed three times for five to ten minutes each with PBS containing 0.1-0.5% Tween-20. Presence of the primary antibody was detected using either an Enhanced Chemiluminescent (ECL) horseradish peroxidase (HRP)-linked secondary antibody or a Near Infrared (NIR) fluorophore-linked antibody. ECL anti-mouse, anti-rabbit IgG/HRP conjugate (GE Healthcare), goat anti-rabbit IgG-HRP (Santa Cruz Biotechnology), or IRDye® infrared 800CW donkey anti-rabbit IgG (Li-cor) were applied as secondary antibodies at a 1/10,000 dilution for 45 minutes. ECL detection was performed using the Advance Western Blotting Detection Kit (GE Healthcare-Amersham) and exposing filters to Hyperfilm ECL (GE Healthcare-Amersham) at 1-30 second exposure times to obtain the best image. Infrared detection was performed by scanning filters with the Odyssey Infrared Imaging System (Li-cor) and viewing the blot under the 800 nm channel.

3. RESULTS

i. Cloning of MBP-XPG C-terminal Subdomains

Multiple pools of purified plasmid for each protein construct were restriction digested using enzymes EcoRI and HindIII, and separated using Agarose Gel Electrophoresis in comparison to uncut, parental and recombinant pGAZ-MBP vector in order to assess ligation of the XPG C-terminal DNA inserts with the pGAZ-MBP vector, as well as the transformation efficiency into competent DH5α bacterial cells (Figure 7 A and B). All of the bacterial colonies from which the pools of plasmids were purified contained the desired recombinant plasmid, which was noted by the presence of a 6 Kilo-base pair (kbp) band for each pools cut and uncut portions. This 6 kbp running size is the approximated size of the linearized plasmid following complete digestion of both DNA strands by the restriction enzymes, thus the bands appear more densely in the cut portions for each protein construct. Uncut portions displayed two additional bands: a dense, leading 4 kbp band, which is the plasmid in the uncut, circular supercoiled state, and a weak, trailing band beyond the 10 kbp size marker, which is the nicked plasmid, where one strand of DNA was cut during extraction of the plasmid DNA from the bacterial cell. The presence of the linearized and nicked bands in the uncut portions suggests that the DNA was damaged during the plasmid purification from the DH5α cells. However, all of the cut portions displayed an

additional leading band of approximate running size to the XPG C-terminal DNA insert for each protein construct, which validated the ligation of the insert to the pGAZ-MBP vector and confirmed the efficiency of the DNA cloning process.

ii. Recombinant Protein Expression of MBP-XPG C-terminal Subdomains

Induced protein expression and solubility was evaluated for all three proteins by visualization of total protein and soluble extract fractions on Bio-Safe Coomassie stained 12% SDS-PAGE (Figure 8). Several proteins were observed for each constructs fractions, though most of these proteins appeared common to all samples. The over-expression of each protein construct produced distinct bands that were clearly evident and unique to the running weight of the target MBP-XPG C-terminal constructs. MBP-XPG E1 was observed at an approximate running weight of 60 kDa. Both MBP-XPG E2 and MBP-XPG E3 were more pronouncedly over-expressed and appeared at approximate running weights of 62 and 65 kDa, respectively. All three recombinant proteins were abundantly over-expressed in E. coli and resulted in stable and soluble recombinant proteins.

iii. Affinity Purification of MBP-XPG C-terminal Subdomains

Protein purity and abundance were determined by Bio-Safe Coomassie stained SDS-PAGE after large scale protein purification of each construct was carried out (Figure 9 A). Binding of the polyhistidine-tagged MBP-XPG C-terminal constructs appeared both efficient and specific as noted by the homogeneous abundance of the target proteins in the eluted portions. Contaminating proteins were observed in the flow-through portion, due to their inability to bind to HisPur Cobalt or Ni-NTA resin. Each protein construct eluted in similar fashion as observed by the noticeable increase in band density over fractions five through nine. Peak elution fractions were pooled and concentrated in storage buffer to 1.6 mg/ mL for MBP-XPG E1, 64 mg/ mL for MBP-XPG E2, and 7.9 mg/ mL for MBP-XPG E3. Dilutions were then made using storage buffer in order to obtain concentrations reasonable for use in biochemical assays (Figure 9 B). MBP-XPG E2 was diluted twenty-fold to a concentration of 3.2 mg/ mL and MBP-XPG E3 was diluted ten-fold to a concentration of 0.8 mg/ mL. All three XPG C-terminal constructs were purified to near homogeneity and concentrated to reasonable quantities.

iv. Protein-Protein Interaction Assay

Previous studies have shown that the extended XPG C-terminus is involved in interaction with a multitude of protein partners. Although the roles of these interactions among the proteins used in this study have been examined to varying degrees *in vitro*, no precise biological roles for these interactions have been determined *in vivo* due to the lack of information regarding the binding site. With the ultimate goal of identifying the residues important for interaction with seven proteins that represent different DNA metabolic pathways, extensive protein-protein interaction mapping was carried out via Far Western analysis using the three sequential XPG C-terminal subdomains E1, E2, and E3. Full length XPG and MBP-XPG Exon-15, which comprises all 180 residues of the C-terminus, were also used in the analysis to serve as positive controls and to provide a comparison interaction by which to assess the affinity of the protein partners for the XPG C-terminal subdomains. Since the XPG C-terminal domain and subdomains were fused to MBP to aid in folding and stabilize protein expression, MBP was additionally included in the mapping studies to serve as a negative control. The four DNA glycosylase proteins being used in this mapping study (NEIL1, NEIL2, NTH1, and APE1), were also included as positive controls on the corresponding Far Western analysis for efficiency of Western blotting with antibodies. Definitive interactions should appear for a protein and its corresponding antibody. Each of these polypeptides (full length XPG, XPG constructs, MBP, and additional controls) were resolved by SDS-PAGE, immobilized to a nitrocellulose filter, and stained for total protein visibility (Figure 10).

When the renatured proteins were probed by the seven protein partners chosen for this study, strong interaction bands appeared consistently for full length XPG, MBP-XPG Exon-15, and MBP-XPG E3 on all blots. For all of the interactions tested, there were no interactions observed for MBP-XPG E1 and MBP-XPG E2. Interaction bands were also not observed for MBP with the exception of the αNTH1 blot, which validated the specificity of the interactions observed on the blots. The four additional DNA glycosylase Western positive controls validated interaction specificity in the αNEIL1 and αNEIL2 blots with the observed interaction band at the approximate running weights of these proteins, but called into question the results of the αNTH1 and αAPE1 by the absence of such a band. The failure of both the positive Western control and

negative control proteins in the αNTH1 blot suggested aggregation of the NTH1 protein stock used to probe the blot, due to the non-specificity of the interactions. Absence of an APE1 band in the αAPE1 blot may be due to an inadequate quantity of protein initially loaded onto the gel that is necessary to provide an interaction band. Both of these blots will be repeated using different protein stocks for interaction probing, though the interactions viewed with the XPG constructs are convincing considering their consistency with regard to the results of the five other Far Western blots. Collectively, these results not only confirmed previous findings of the binding of these proteins to full length XPG and the XPG C-terminus, but also determined a narrower 80 amino acid region, comprising residues 1107 to 1186, as being sufficient for interaction with each of the seven proteins used in this study (Figure 11).

4. DISCUSSION AND CONCLUSIONS

The multifunctional roles of XPG in a variety of DNA metabolic pathways are hypothesized to be a result of its ability to interact with a multitude of proteins. This study was able to demonstrate the importance of an eighty amino acid region in the C-terminus for interaction with several proteins from three distinct DNA repair pathways. Based on repeated failed attempts at determining its three-dimensional structure, the C-terminus is thought to be extended and unstructured in the absence of a protein-binding partner. This concept is dually observed in the DAMMIN Ab initio analysis of the molecular envelope that was performed on the MBP-XPG Exon-15 construct, as well as the largely disordered prediction of the C-terminus using Predictor of Natural Disordered Regions (PONDR) analysis (Figure 12 A and B) [32]. An emerging concept in understanding regulation of pleiotropic proteins such as XPG is that conformational changes mediated by interactions with other proteins, DNA, or nucleotides play a significant role in determining their activities and partnerships. In solution, proteins are dynamic flexible molecules and protein-protein interactions are often mediated by intrinsically disordered regions of proteins because flexibility allows the formation of intricate binding surfaces with multiple structurally distinct proteins [33]. Disordered regions can include large stretches of sequence which can assume multiple conformations and may or may not possess intrinsic secondary structure, but which becomes structured upon binding to substrate, posttranslational modification, or interaction with partners. Recent, well-documented examples of this disorder-

to-order transition include the disordered N-terminal tails of histones, the disordered C-terminus of the τ subunit E. coli DNA polymerase III, and the bacterial UmuD protein that regulates translesion synthesis [34]-[36]. A provocative model illustrating the concept that sequential conformational changes in intrinsically disordered domains can control pathway progression has emerged from the study on UmuD protein (Figure 13) [36]. Given the extended and flexible conformation of the E3 region, this concept is supported by the results of this study, which demonstrate the ability for a narrow stretch of eighty amino acids (1107-1186), to mediate interactions with all seven of the DNA metabolic proteins chosen for finer interaction mapping. These seven proteins represent only a portion of the XPG C-terminal binding partners, which suggest an even more versatile ability for E3 to bind proteins than has been demonstrated here.

Though the collective isoelectric point (pI) of 11.3 for E3 suggests that this XPG C-terminal region is basic at physiologic pH, closer examination of the amino acid sequence reveals two distinct regions that fluctuate drastically from this pI and may provide the intrinsic disorder necessary for mediating the multiple protein interactions (Figure 14). The first possible interaction domain appears as the eleven amino acid stretch, ^{1144}TTSSSSDSDDD^{1154} that has a theoretical pI of 3.52 and which contains residues that can potentially form a highly acidic domain via hyperphosphorylated posttranslational activation. Phosphorylation is the addition of a phosphate (PO_4) group to certain amino acid residues capable of such modifications. Reversible phosphorylation of proteins is an important regulatory mechanism that occurs in both prokaryotic and eukaryotic organisms, and usually occurs on serine (S), threonine (T), and tyrosine (Y) residues in eukaryotic proteins. The addition of a PO_4 molecule to a polar R group of an amino acid residue can turn a hydrophobic portion of a protein into a polar and extremely hydrophilic portion of molecule. In this way it can introduce a conformational change in the structure of the protein via interaction with other hydrophobic and hydrophilic residues in the protein. The four negatively charged aspartate (D) residues that are grouped in this region may provide necessary interaction surfaces with the seven S and T amino acids to form a highly acidic eleven amino acid residue domain that is capable of mediating interactions with multiple protein partners.

The second proposed interaction domain may be the fifteen amino acid stretch, ^{1171}KKRRKLRRARGRKRK^{1185} that has an extremely basic theoretical pI of 12.78 due to the predominance of negatively charged amino acids, arginine (R) and lysine (K). This exact sequence has also been established as the nuclear localization signal (NLS) for XPG, which is the amino acid sequence 'tag' on the proteins surface that regulates the dynamic localization of the protein in the cell nucleus [37]. Although other candidate NLS peptides have been proposed for XPG protein, this particular NLS is conserved evolutionarily between yeasts and humans, and has been shown to direct XPG protein to the nucleus following UV-irradiation of the cell [37]. These properties make it difficult to determine if the same amino acid region is additionally capable of mediating multiple interactions with other proteins. However, the domain provides a stretch of exposed, hydrophilic residues that presumable possess no intrinsic secondary structure, which are necessary for binding of multiple proteins. If protein-protein interactions are found to be mediated by this region in future mapping studies, conducting biochemistry experiments to study the biological significance of each protein-protein interaction without interrupting the NLS functions would be very challenging. NLS interruption would impede the ability to determine the isolated phenotypic consequences of disrupting single protein-protein interactions, *in vivo*.

Since all of the protein partners used for the fine-mapping are interacting with the same C-terminal subdomain, additional studies are required to determine if these interactions are mutually exclusive within the E3 regions eighty amino acids and if so, identify the precise interaction signatures of each of these proteins. To accomplish this goal, the Cooper lab will use peptide arrays covering the E3 region of XPG to screen for interacting residues with each of the C-terminal binding partners. Each array will be synthesized with overlapping 12-mer peptides offset by two residues. ABIMED cellulose filter peptide arrays will be purchased from the Biopolymers Laboratory at MIT Cancer Center through a collaborator, Graham Walker, whose lab has recently used such arrays to finely map the interaction between bacterial DinB and UmuD2 [38]. Filters can contain up to 384 peptide spots and can be re-used ten to twenty times. Filters will be probed with each of the C-terminal interacting partners, followed by western blot detection using an antibody against the interacting protein. Based on the mapping results,

"surgical" deletion constructs will be designed in which precise interacting amino acid residues will be removed. The peptide scan may reveal overlap in the E3 regions that mediate interactions with C-terminal binding partners. Therefore, alanine mutagenesis will additionally be used within this region to more finely map and disrupt protein interaction sites. Recently this method was used to reveal distinct interaction sites for five proteins in the C-terminal domain of the XPG homolog, FEN-1 [39]. Once specific sites of interaction are identified, mutant proteins that are defective only in the deletion or substitution of amino acids required for a certain protein-protein interaction will be designed for both full-length XPG and the XPG C-terminus. Inability of the mutant proteins to interact with the protein requiring the mutated amino acids will be verified, as well as the ability of the mutant protein to preserve interactions with other partner proteins. Ultimately these studies intend to elucidate the phenotypic consequences of disrupting certain XPG protein-protein interactions by examining the mutant proteins biochemical potentials via enzymatic, DNA binding, and annealing activities, as well as ability to stimulate certain TCR, BER and DSBR proteins.

Evidently, many of the cellular roles for XPG protein remain to be defined. This study has provided notable insights into XPG protein function, gained from methods aimed at defining XPG interactions with partners. Since XPG is often required to form larger multi-component cellular machines, with conformational changes and protein hand-offs likely controlling TCR pathway regulation, the continuation of these studies will likely provide a detailed understanding of how mutation of a specific gene results in the complex XP-CS phenotype and ultimately how these DNA repair processes maintain the integrity of the human genome.

5. ACKNOWLEDGEMENTS

This research was conducted at the Cooper Lab of Lawrence Berkeley National Laboratory for the Department of Energy, 2009 Science Undergraduate Laboratory Internship summer program, and was supported by the Center for Science and Engineering Education and the DOE Office of Science. I would like to thank my mentor, Dr. Kelly S. Trego, for her amazing mentorship and endless guidance. Thanks also to Dr. Priscilla K. Cooper, Miaw-Sheue Tsai, Altaf Sarker, Cliff Ng, Susan Tsutakawa, Gareth Williams, Jill Fuss, Claudia Wiese, Doina Ciobanu, Torsten

Groesser, Noah Schwartz and the CSEE staff for assisting various aspects of this research. Lastly, thanks to LBNL and the DOE Office of Science for creating a truly fulfilling research program, and for hosting me and funding my research over two fantastic summers.

6. REFERENCES

[1] D.P. Batty, R.D. Wood, "Damage recognition in nucleotide excision repair of DNA," Gene, vol. 241, no. 22, 193-204, 2000.

[2] J.O. Fuss, P.K. Cooper, "DNA repair: dynamic defenders against cancer and aging," Public Library of Science Biology, vol. 4, no. 6, 899-903, 2006.

[3] P.C. Hanawalt, "Subpathways of nucleotide excision repair and their regulation," Oncogene, vol. 21, no. 58, 8949-8956, 2002.

[4] I. Mellon, G. Spivak, P.C. Hanawalt, "Selective removal of transcription-blocking DNA damage from the transcribed strand of the mammalian DHFR gene," Cell, vol. 51, no. 2, 241-249, 1987.

[5] M. Ljungman, D.P. Lane, "Transcription—guarding the genome by sensing DNA damage," Nature Reviews Cancer, vol. 4, no. 9, 727-737, 2004.

[6] A.R. Lehmann, "DNA repair-deficient diseases, xeroderma pigmentosum, Cockayne syndrome and trichothiodystrophy," Biochimie, vol. 85, no. 11, 1101-1111, 2003.

[7] K.H. Kraemer, N.J. Patronas, R. Schiffmann, B.P. Brooks, D. Tamura, J.J. DiGiovanna, "Xeroderma pigmentosum, trichothiodystrophy and Cockayne syndrome: a complex genotype-phenotype relationship," Neuroscience, vol. 145, no. 4, 1388-1396, 2007.

[8] Y. Lindenbaum, D. Dickson, P. Rosenbaum, K. Kraemer, I. Robbins, I. Rapin, "Xeroderma pigmentosum/cockayne syndrome complex: first neuropathological study and review of eight other cases," European Journal of Paediatric Neurology, vol. 5, 225-242, 2001.

[9] H. de Waard, J. de Wit, J.O. Andressoo, C.T. van Oostrom, B. Riis, A. Weimann, H.E. Poulsen, H. van Steeg, J.H. Hoeijmakers, G.T. van der Horst, "Different effects of CSA and CSB deficiency on sensitivity to oxidative DNA damage," Molecular and Cellular Biology, vol. 24, no. 18, 7941-7948, 2004.

[10] G. Spivak, P.C. Hanawalt, "Host cell reactivation of plasmids containing oxidative DNA lesions is defective in Cockayne syndrome but normal in UV-sensitive syndrome fibroblasts," DNA Repair (Amsterdam), vol. 5, no. 1, 13-22, 2006.

[11] M. D'Errico, E. Parlanti, M. Teson, P. Degan, T. Lemma, A. Calcagnile, I. Iavarone, P. Jaruga, M. Ropolo, A.M. Pedrini, D. Orioli, G. Frosina, G. Zambruno, M. Dizdaroglu, M. Stefanini, E. Dogliotti, "The role of CSA in the response to oxidative DNA damage in human cells," Oncogene, vol. 26, no. 30, 4336-4344, 2007.

[12] A. Klungland, M. Hoss, D. Gunz, A. Constantinou, S.G. Clarkson, P.W. Doetsch, P.H. Bolton, R.D. Wood RD, T. Lindahl, "Base excision repair of oxidative DNA damage activated by XPG protein," Molecular Cell, vol. 3, no. 1, 33-42, 1999.

[13] T. Bessho, "Nucleotide excision repair 3′ endonuclease XPG stimulates the activity of base excision repair enzyme thymine glycol DNA glycosylase," Nucleic Acids Research, vol. 27, no. 4, 979-983, 1999.

[14] S.K. Lee, S.L. Yu, L. Prakash, S. Prakash, "Requirement of yeast RAD2, a homolog of human XPG gene, for efficient RNA polymerase II transcription. Implications for Cockayne syndrome," Cell, vol. 109, no. 7, 823-834, 2002.

[15] D. Scherly, T. Nouspikel, J. Corlet, C. Ucla, A. Bairoch, S.G. Clarkson, "Complementation of the DNA repair defect in xeroderma pigmentosum group G cells by a human cDNA related to yeast RAD2," Nature, vol. 363, no. 6425, 182-185, 1993.

[16] A. O'Donovan, R.D. Wood, "Identical defects in DNA repair in xeroderma pigmentosum group G and rodent ERCC group 5," Nature, vol. 363, no. 6425, 185-188, 1993.

[17] A. O'Donovan, A.A. Davies, J.G. Moggs, S.C. West, R.D. Wood, "XPG endonuclease makes the 3′ incision in human DNA nucleotide excision repair," Nature, vol. 371, no. 6496, 432-435, 1994.

[18] T.A. Ceska, J.R. Sayers, G. Stier, D. Suck, "A helical arch allowing single-stranded DNA to thread through T5 5′-exonuclease," Nature, vol. 382, no. 6586, 90-93, 1996.

[19] D.J. Hosfield, C.D. Mol, B. Shen, J.A. Tainer, "Structure of the DNA repair and replication endonuclease and exonuclease FEN-1: coupling DNA and PCNA binding to FEN-1 activity," Cell, vol. 95, no. 1, 135-146, 1998.

[20] T.C. Mueser, N.G. Nossal, C.C. Hyde, "Structure of bacteriophage T4 RNase H, a 5′ to 3′ RNA-DNA and DNA-DNA exonuclease with sequence similarity to the RAD2 family of eukaryotic proteins," Cell, vol. 85, no. 7, 1101-1112, 1996.

[21] K.Y. Hwang, K. Baek, H.Y. Kim, Y. Cho, "The crystal structure of flap endonuclease-1 from Methanococcus jannaschii," Nature Structural and Molecular Biology, vol. 5, no. 8, 707-713, 1998.

[22] S. Emmert, H. Slor, D.B. Busch, S. Batko, R.B. Albert, D. Coleman, S.G. Khan, B. Abu-Libdeh, J.J. DiGiovanna, B.B. Cunningham, M.M Lee, J. Crollick, H. Inui, T. Ueda, M. Hedayati, L. Grossman, T. Shahlavi, J.E. Cleaver, K.H. Kraemer, "Relationship of neurologic degeneration to genotype in three xeroderma pigmentosum group G patients," Journal of Investigative Dermatology, vol. 118, no. 6, 972-982, 2002.

[23] P. Lalle, T. Nouspikel, A. Constantinou, F. Thorel, S.G. Clarkson, "The founding members of xeroderma pigmentosum group G produce XPG protein with severely impaired endonuclease activity," Journal of Investigative Dermatology, vol. 118, no. 2, 344-351, 2002.

[24] T. Nouspikel, P. Lalle, S.A. Leadon, P.K. Cooper, S.G. Clarkson, "A common mutational pattern in Cockayne syndrome patients from xeroderma pigmentosum group G: implications for a second XPG function," Proceedings of the National Academy of Sciences of the United States of America, vol. 94, no. 7, 3116-3121, 1997.

[25] F. Thorel, A. Constantinou, I. Dunand-Sauthier, T. Nouspikel, P. Lalle, A. Raams, N.G. Jaspers, W. Vermeulen, M.K. Shivji, R.D. Wood, S.G. Clarkson, "Definition of a short

region of XPG necessary for TFIIH interaction and stable recruitment to sites of UV damage," Molecular and Cellular Biology, vol. 24, no. 24, 10670-10680, 2004.

[26] I. Dunand-Sauthier, M. Hohl, F. Thorel, P. Jaquier-Gubler, S.G. Clarkson, O.D. Schärer, "The spacer region of XPG mediates recruitment to nucleotide excision repair complexes and determines substrate specificity," Journal of Biological Chemistry, vol. 280, no. 8, 7030-7037, 2005.

[27] M. Hohl, I. Dunand-Sauthier, L. Staresincic, P. Jaquier-Gubler, F. Thorel, M. Modesti, S.G. Clarkson, O.D. Schärer, "Domain swapping between FEN-1 and XPG defines regions in XPG that mediate nucleotide excision repair activity and substrate specificity," Nucleic Acids Research, vol. 35, no. 9, 3053-3063, 2007.

[28] A.H. Sarker, S.E. Tsutakawa, S. Kostek, C. Ng, D.S. Shin, M. Peris, E. Campeau, J.A. Tainer, E. Nogales, P.K. Cooper, "Recognition of RNA polymerase II and transcription bubbles by XPG, CSB and TFIIH: insights for transcription-coupled repair and Cockayne syndrome," Molecular Cell, vol. 20, no. 2, 187-198, 2005.

[29] S. Ito, I. Kuraoka, P. Chymkowitch, E. Compe, A. Takedachi, C. Ishigami, F. Coin, J.M. Egly, K. Tanaka, "XPG stabilizes TFIIH, allowing transactivation of nuclear receptors: implications for Cockayne syndrome in XP-G/CS patients," Molecular Cell, vol. 26, no. 2, 231-243, 2007.

[30] H. Dou, S. Mitra, T.K. Hazra, "Repair of oxidized bases in DNA bubble structures by human DNA glycosylases NEIL1 and NEIL2," Journal of Biological Chemistry, vol. 278, no. 50, 49679-49684, 2003.

[31] J.J.P. Perry, S.M. Yannone, L.G. Holden, C. Hitomi, A. Asaithamby, S. Han, P.K. Cooper, D.J. Chen, J.A. Tainer, "WRN exonuclease structure and molecular mechanism imply an editing role in DNA end processing," Nature Structural and Molecular Biology, vol. 13, no. 5, 414-422, 2006.

[32] X. Li, P. Romero, M. Rani, A.K. Dunker, Z. Obradovic, "Predicting protein disorder for N-, C- and internal regions," Genome Informatics Series Workshop, vol. 10, 30-40, 1999.

[33] H.J. Dyson, P.E. Wright, "Coupling of folding and binding for unstructured proteins," Current Opinion in Structural Biology, vol. 12, 54-60, 2002.

[34] S.L. Berger, "Histone modifications in transcriptional regulation," Current Opinion in Genetics and Development, vol. 12, 142-148, 2002.

[35] S. Jergic, K. Ozawa, N.K. Williams, X.C. Su, D.D. Scott, S.M. Hamdan, J.A. Crowther, G. Otting, N.E. Dixon, "The unstructured C-terminus of the tau subunit of Escherichia coli DNA polymerase III holoenzyme is the site of interaction with the alpha subunit," Nucleic Acids Research, vol. 35, 2813–2824, 2007.

[36] S.M. Simon, F.J. Sousa, R. Mohana-Borges, G.C. Walker, "Regulation of Escherichia coli SOS mutagenesis by dimeric intrinsically disordered umuD gene products," Proceedings of the National Academy of Sciences of the United States of America, vol. 105, 1152-1157, 2008.

[37] J.A. Knauf, S.H. Pendergrass, B.L. Marrone, G.F. Strniste, M.A. MacInnes, M.S. Park, "Multiple nuclear localization signals in XPG nuclease," Mutation Research, vol. 363, no. 1, 67-75, 1996.

[38] V.G. Godoy, D.F. Jarosz, S.M. Simon, A. Abyzov, V. Ilyin, G.C. Walker, "UmuD and RecA directly modulate the mutagenic potential of the Y family DNA polymerase DinB," Molecular Cell, vol. 28, 1058-1070, 2007.

[39] Z. Guo, V. Chavez, P. Singh, L.D. Finger, H. Hang, M.L. Hegde, B. Shen, "Comprehensive mapping of the C-terminus of flap endonuclease-1 reveals distinct interaction sites for five proteins that represent different DNA replication and repair pathways," Journal of Molecular Biology, vol. 377, 679-690, 2008.

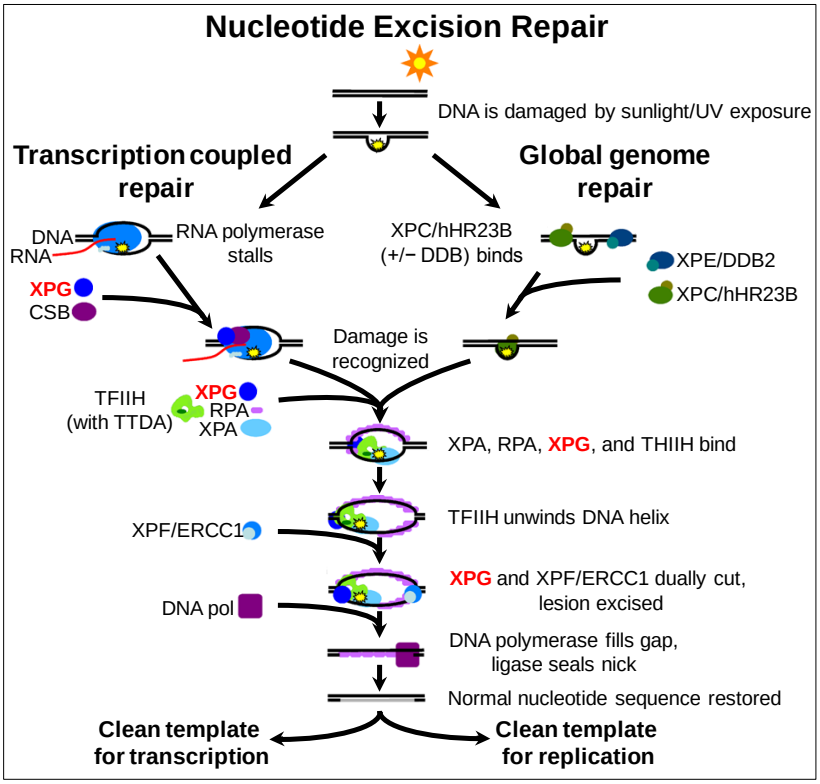

Figure 1. Nucleotide Excision Repair schematic. When DNA is damaged by sunlight, the damage is recognized differently depending on whether the DNA is transcriptionally active (transcription-coupled repair) or not (global genome repair). After the initial recognition step, the damage is repaired in similar manner with the final outcome being the restoration of the normal nucleotide sequence [2].

Figure 2. Patient XP20BE with XP-CS complex at age 6 years. Cockayne syndrome sufferers have multiple systemic disorders due to a defect in the ability of cells to repair DNA that is being transcribed. Signs of advanced CS are observed by the profound weakness of dorsiflexion of the hand in addition to the short stature in comparison to the mother who is holding him [8].

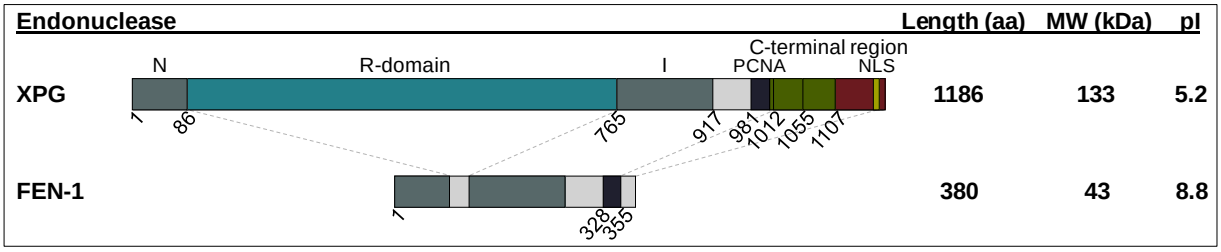

Figure 3. XPG and FEN-1 schematic. Though both XPG and FEN-1 are structure specific endonucleases, XPG protein is considerably larger than FEN-1 due to the uniquely extended R- and C-terminus domains, which impart XPG with additional functional capabilities. Extensive sequence homology is conserved between the catalytic sub-domains N and I of XPG and FEN-1. Comparative differences in terms of polypeptide length, molecular weight and isoelectric point are shown to the right of each protein. XPG is considerably more acidic than FEN-1.

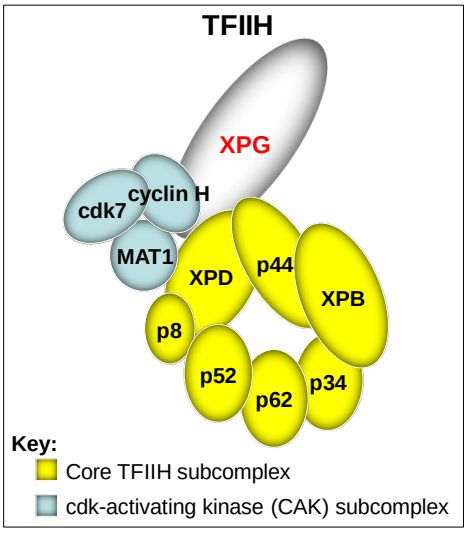

Figure 4. Schematic model for the role of XPG in maintaining the integrity of TFIIH. Wild-type XPG forms a complex with TFIIH and functions in maintaining the integrity of the two subcomplexes. Mutant XPG with a C-terminal deletion that is derived from a XPG-CS patient causes the dissociation of CAK and XPD from the core TFIIH (not shown) [29].

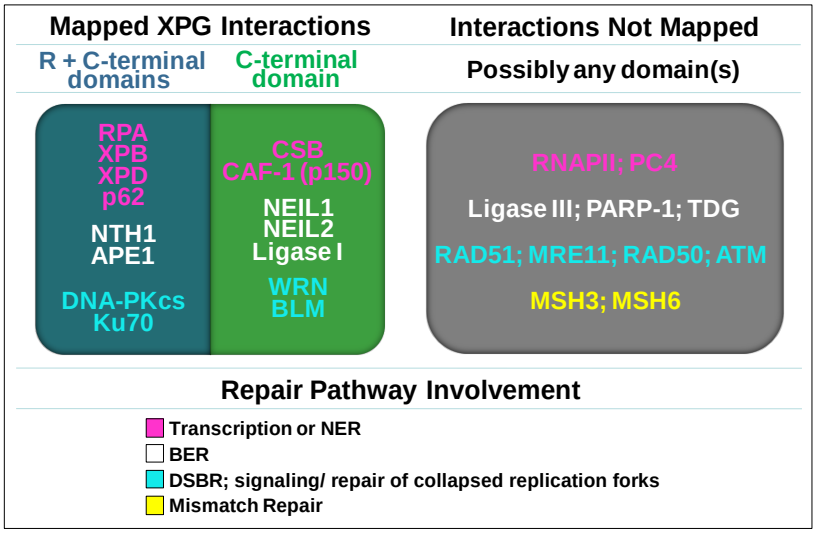

Figure 5. Diagram of XPG interacting proteins. XPG interacts with proteins involved in NER/Transcription (magenta), BER (white), and DSBR (sky blue). Interactions are mediated by both of the largely unstructured R- and C-terminal domains (teal box) and by the C-terminal domain alone (green box). Multiple unmapped binding partners of XPG exist (grey box) that are also involved in these repair pathways and mismatch repair (yellow).

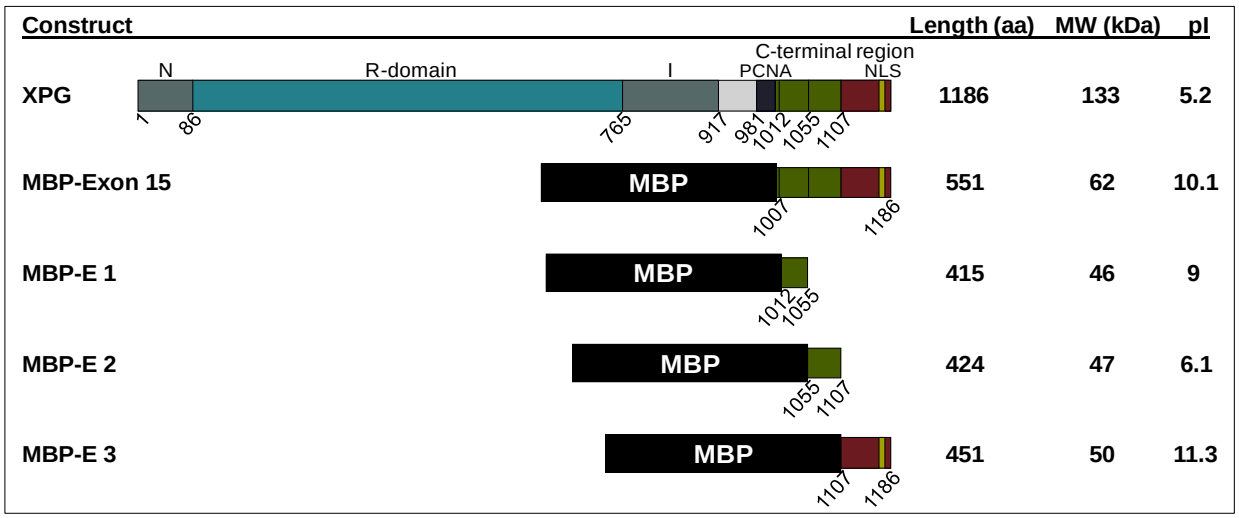

Figure 6. Schematic diagram of the subdomain constructs of the XPG C-terminus in relation to full length XPG and the XPG C-terminus (Exon 15). XPG C-terminus was sub-cloned into three fragments based upon PONDR prediction of folding, as shown in figure 12 B. The N-terminus of each subdomain was fused to the C-terminus of the 42-3 kDa (380 amino acid) MBP (shown in black), and was expressed and purified in E. coli. MBP fusion aided in the folding, stable expression, and affinity purification of the constructs. Comparative statistics of XPG, Exon 15, and the three subdomains are displayed to the right of each protein. The numbers *below* each construct represent amino acid positions within full length XPG protein. Together, all five of these proteins were used in the interaction mapping with seven C-terminal binding protein partners.

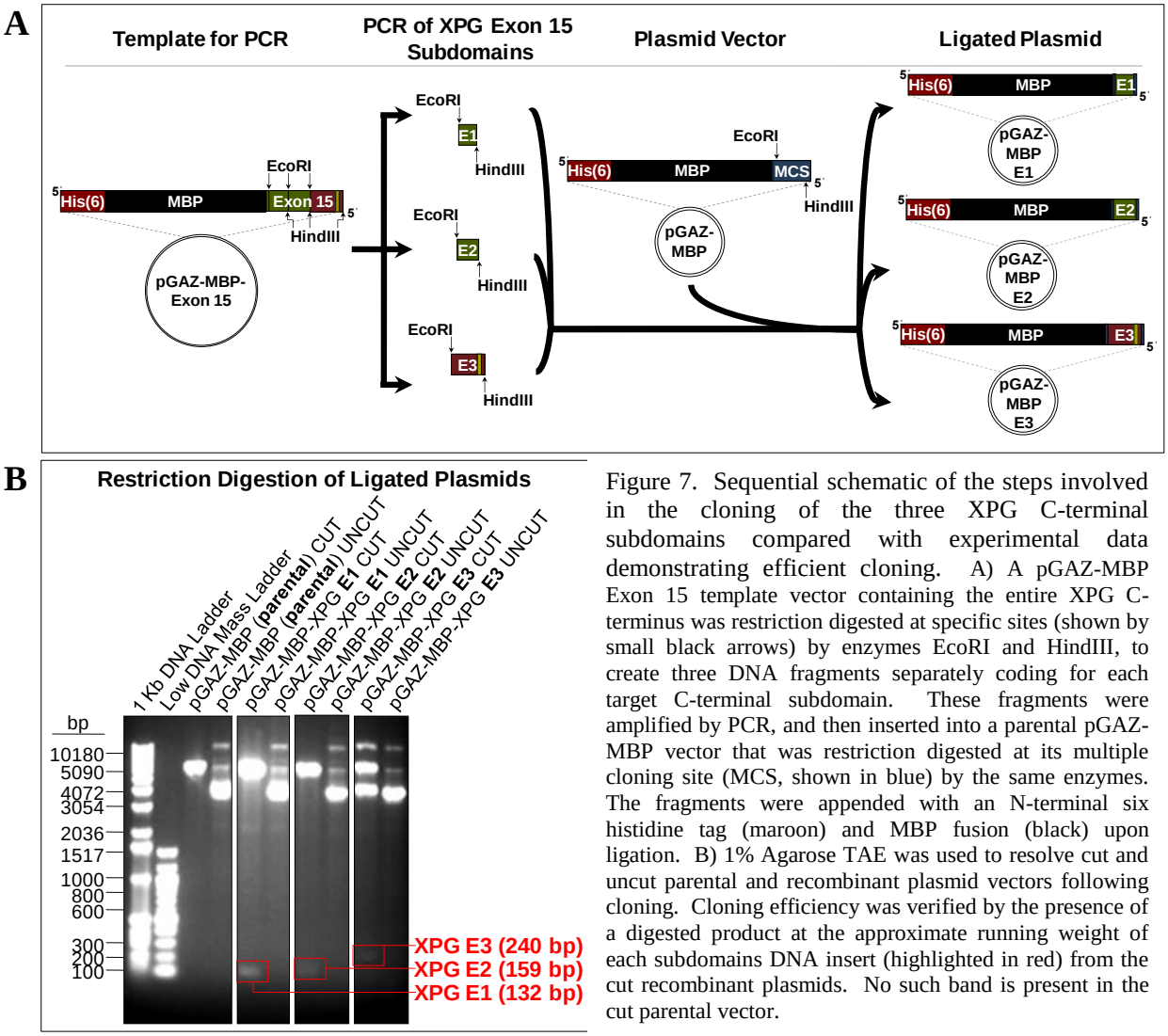

Figure 7. Sequential schematic of the steps involved in the cloning of the three XPG C-terminal subdomains compared with experimental data demonstrating efficient cloning. A) A pGAZ-MBP Exon 15 template vector containing the entire XPG C-terminus was restriction digested at specific sites (shown by small black arrows) by enzymes EcoRI and HindIII, to create three DNA fragments separately coding for each target C-terminal subdomain. These fragments were amplified by PCR, and then inserted into a parental pGAZ-MBP vector that was restriction digested at its multiple cloning site (MCS, shown in blue) by the same enzymes. The fragments were appended with an N-terminal six histidine tag (maroon) and MBP fusion (black) upon ligation. B) 1% Agarose TAE was used to resolve cut and uncut parental and recombinant plasmid vectors following cloning. Cloning efficiency was verified by the presence of a digested product at the approximate running weight of each subdomains DNA insert (highlighted in red) from the cut recombinant plasmids. No such band is present in the cut parental vector.

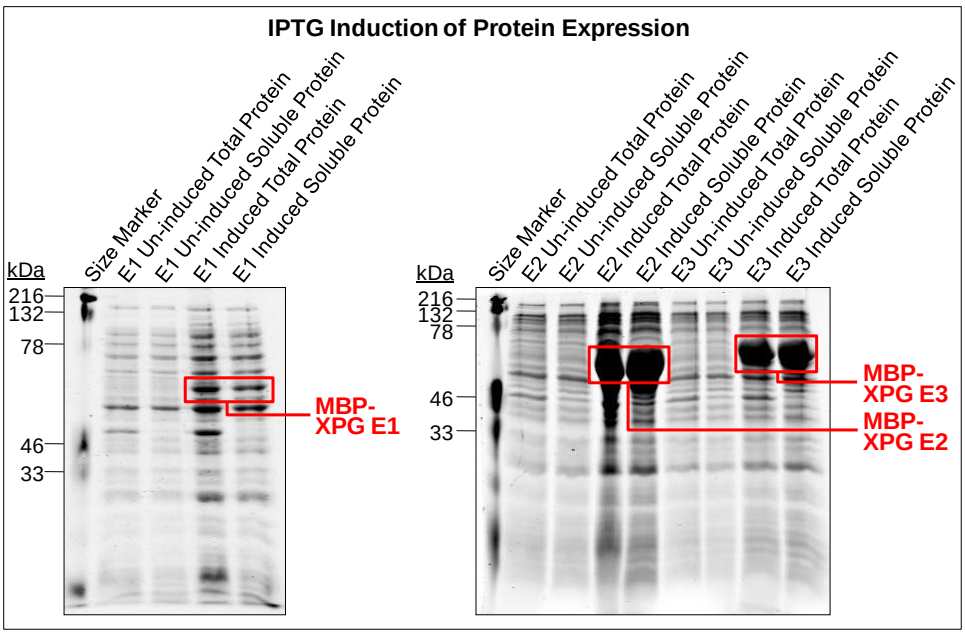

Figure 8. Resolved upon 12% SDS-PAGE, un-induced and induced total and soluble fractions for each C-terminal subdomain reveals the presence of entirely soluble target proteins (highlighted in red) following IPTG induction. This demonstrates the efficiency of the protein expression.

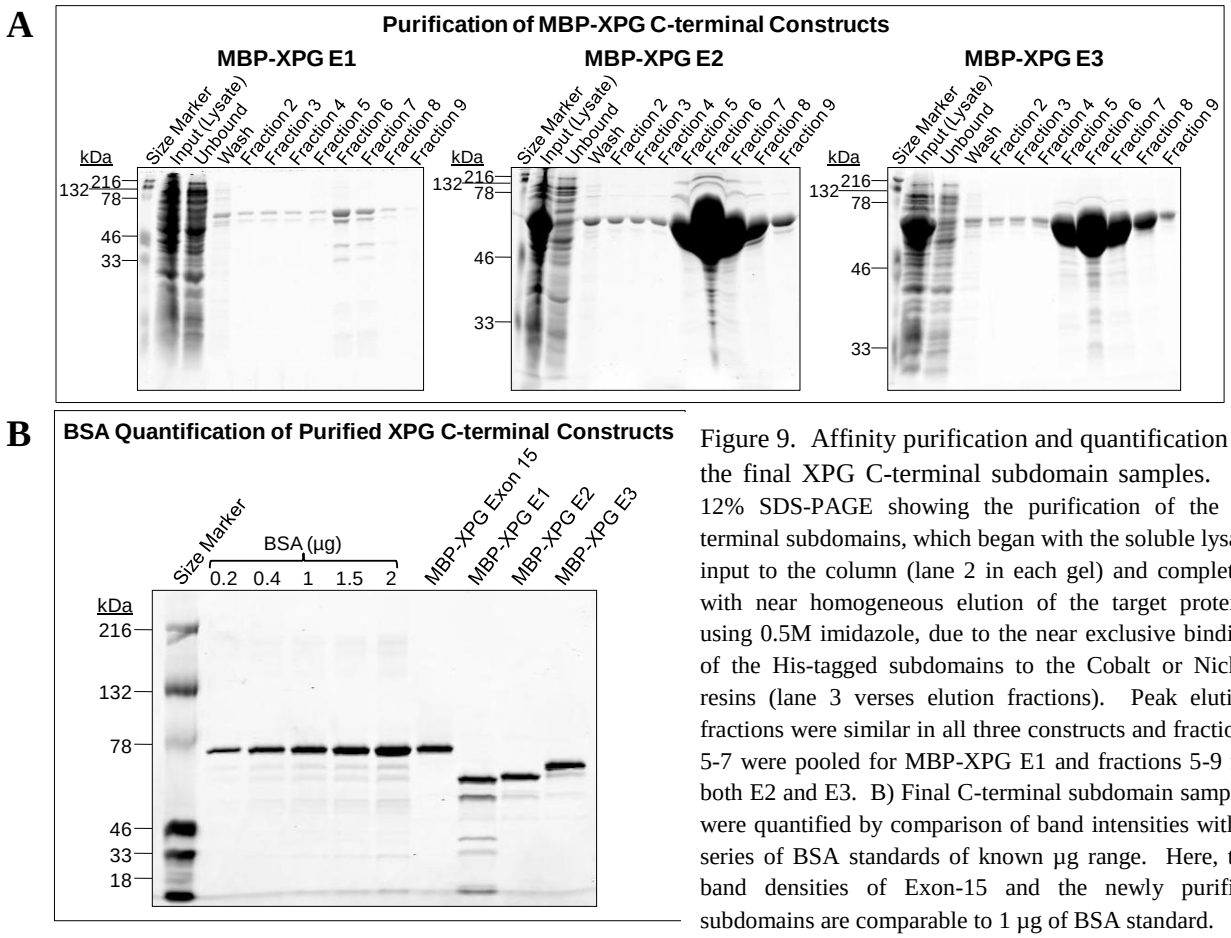

Figure 9. Affinity purification and quantification of the final XPG C-terminal subdomain samples. A) 12% SDS-PAGE showing the purification of the C-terminal subdomains, which began with the soluble lysate input to the column (lane 2 in each gel) and completed with near homogeneous elution of the target proteins using 0.5M imidazole, due to the near exclusive binding of the His-tagged subdomains to the Cobalt or Nickel resins (lane 3 verses elution fractions). Peak elution fractions were similar in all three constructs and fractions 5-7 were pooled for MBP-XPG E1 and fractions 5-9 for both E2 and E3. B) Final C-terminal subdomain samples were quantified by comparison of band intensities with a series of BSA standards of known µg range. Here, the band densities of Exon-15 and the newly purified subdomains are comparable to 1 µg of BSA standard.

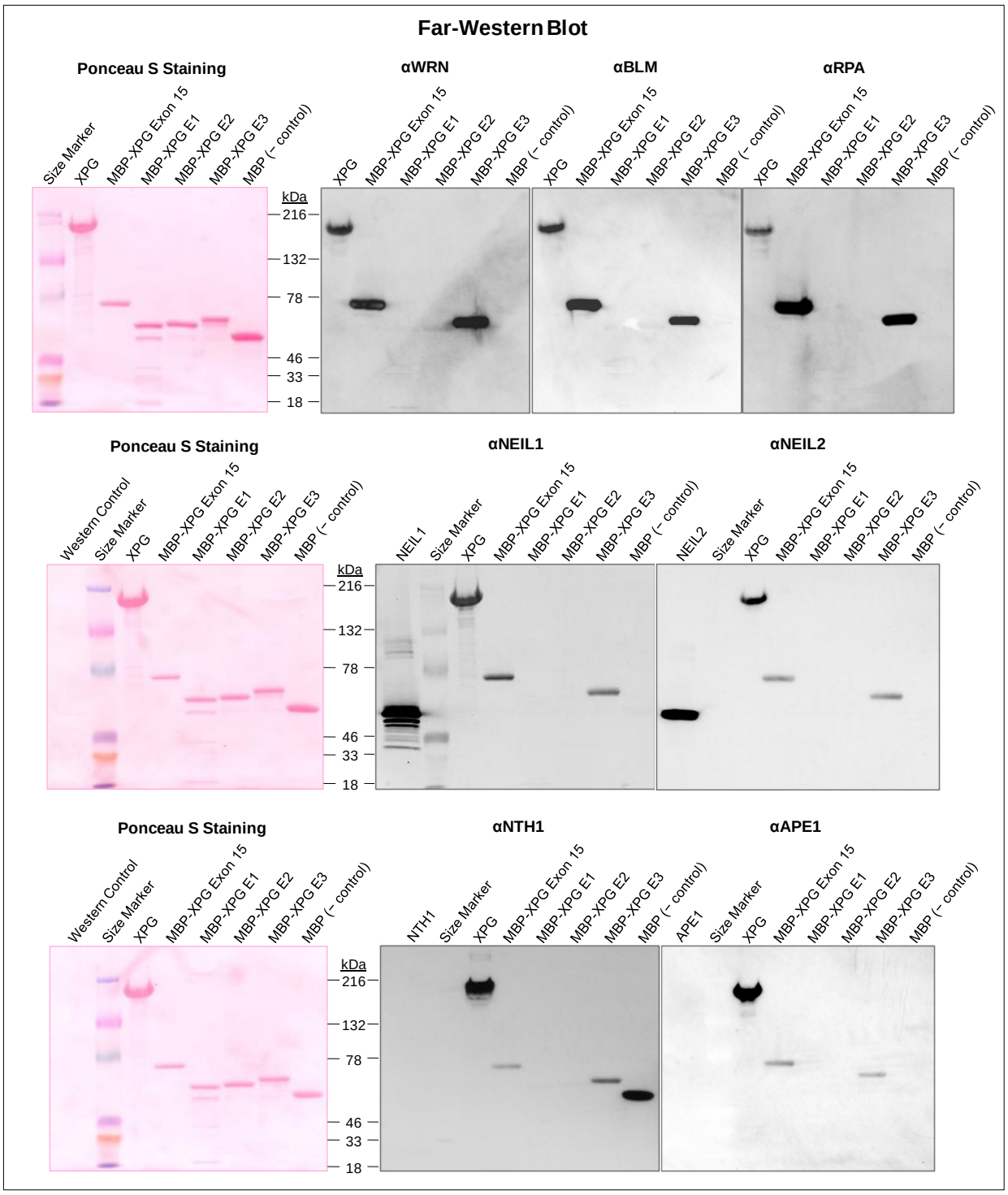

Figure 10. Mapping of interaction regions of XPG with seven proteins by Far Western assay. Full length XPG, Exon 15, and the 3 C-terminal subdomains (E1, E2, and E3), as well as the negative control MBP were resolved by 4-12% Tris-glycine SDS-PAGE in the order just described. Additional Western positive controls were used in the four glycosylase blots and were resolved in lane 1. The resolved proteins in each blot were immobilized to nitrocellulose, which was then temporarily stained with Ponceau S solution for total protein visibility (left side of each row). Filters were then blocked and incubated with one of the seven C-terminal binding proteins, and immunodetected with antibody against the binding partner, resulting in the seven blots which appear in black and white. Protein-protein interactions were observed by bands, which consistently occurred for full length XPG, Exon 15, and E3 in all of the blots.

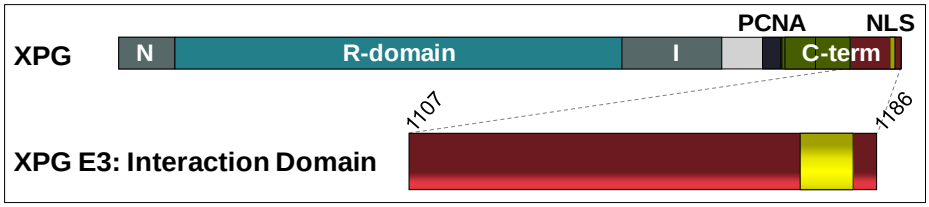

Figure 11. Schematic representation of full length XPG in comparison to the 80 amino acid region of the C-terminus that mediates interactions with all seven of the protein partners examined by Far Western analysis. Though this region is narrow in comparison to the full length protein, the multiple protein-protein interactions are likely mediated by a few amino acids or a certain motif, which would require additional studies to identify precisely.

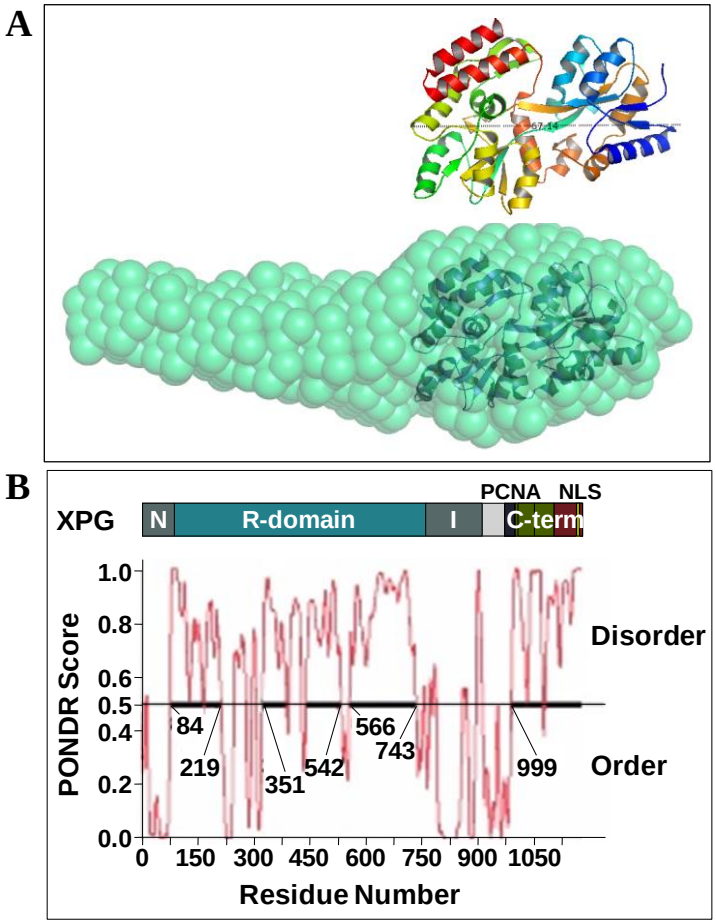

Figure 12. A) Dammin ab initio shape reconstruction of MBP-XPG C-terminus with crystal structure of MBP (top) superimposed, demonstrating the tremendous extension of the C-terminus. B) PONDR diagram of XPG showing areas of order and disorder. PONDR scores of 0.5 or higher represent regions of disorder. Alignment of an XPG schematic with the PONDR plot demonstrates the increased disorderliness in the R- and C-terminal domains.

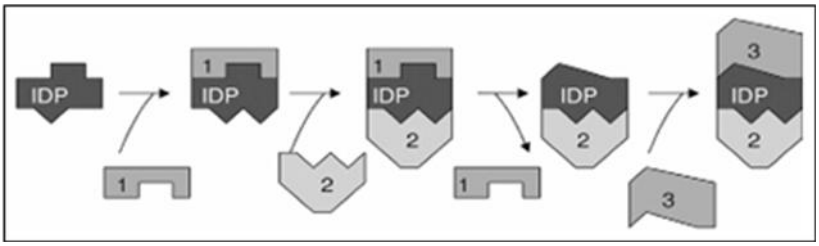

Figure 13. Model for sequential protein-protein interactions with an Intrinsically Disordered Protein (IDP). An IDP may first bind to one interaction patner (1), which stabilizes a particular conformation. If a second binding interface becomes exposed in this conformation, another protein may now bind (2). The second binding event could destabilize the first protein-protein interaction, causing the original protein to exit the complex and possibly exposing a different interface. If so, a different partner (3) can bind at this site [36].

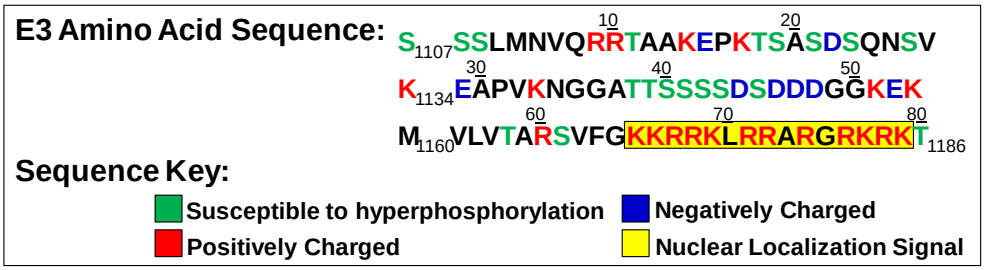

Figure 14. Eighty amino acid sequence of interacting XPG C-terminal subdomain. Charged residues, as well as residues capable of hyperphosphorylated mediation are designated by the color scheme described in the key. Drastic polarity shift in the E3 region is evident in the extremely acidic 1144-1154 stretch (pI 3.52) being separated by only 18 residues from the extremely basic 1171-1185 stretch (pI 12.78). The later stretch (highlighted in yellow), is also the NLS domain.