Biochemical and Structural Domain Analysis of Xeroderma Pigmentosum Complementation Group C Protein†

Christopher G. Bunick,‡ Michael R. Miller,‡,§ Brian E. Fuller,‡,§ Ellen Fanning,§ and Walter J. Chazin*,‡

Departments of Biochemistry and Chemistry and Center for Structural Biology, 5140 BIOSCI/MRBIII, and Department of Biological Sciences, Vanderbilt University, Nashville, Tennessee 37232-8725

Received July 6, 2006; Revised Manuscript Received September 8, 2006

ABSTRACT: XPC is a 940-residue multidomain protein critical for the sensing of aberrant DNA and initiation of global genome nucleotide excision repair. The C-terminal portion of XPC (residues 492−940; XPC-C) has critical interactions with DNA, RAD23B, CETN2, and TFIIH, whereas functional roles have not yet been assigned to the N-terminal portion (residues 1−491; XPC-N). In order to analyze the molecular basis for XPC function and mutational defects associated with xeroderma pigmentosum (XP) disease, a series of stable bacterially expressed N- and C-terminal fragments were designed on the basis of sequence analysis and produced for biochemical characterization. Limited proteolysis experiments combined with mass spectrometry revealed that the full XPC-C is stable but XPC-N is not. However, a previously unrecognized folded helical structural domain was found within XPC-N, XPC(156−325). Pull-down and protease protection assays demonstrated that XPC(156−325) physically interacts with the DNA repair factor XPA, establishing the first functional role for XPC-N. XPC-C exhibits binding characteristics of the full-length protein, including stimulation of DNA binding by physical interaction with RAD23B and CETN2. Analysis of an XPC missense mutation (Trp690Ser) found in certain patients with XP disease revealed that this mutation is associated with a diminished ability to bind DNA. Evidence of contributions to protein interactions from regions in both XPC-N and XPC-C along with recently recognized homologies to yeast PNGase prompted construction of a structural model of a folded XPC core. This model offers key insights into how domains from the two portions of the protein may cooperate in generating specific XPC functions.

Xeroderma pigmentosum complementation group C (XPC)[1] is a 940 amino acid protein vital for sensing the presence of aberrant DNA and initiating global genome nucleotide excision DNA repair (GG-NER) (*1*). XPC recognizes a broad spectrum of helix-distorting lesions present in transcriptionally inactive DNA (*2*−*6*), including bulky ultraviolet radiation-induced lesions (6−4 photoproducts and cyclobutane−pyrimidine dimers) as well as cisplatin adducts (*7*−*9*). Damage recognition by XPC is stimulated through its protein−protein interactions with the ubiquitin-like protein human homologue of Rad23 B (RAD23B) (*10*) and the EF-hand calcium binding protein centrin-2 (CETN2) (*11*). The heterotrimeric XPC−RAD23B−CETN2 complex participates in the first fundamental step of GG-NER, DNA damage recognition, and sets the stage for the following two steps, DNA damage excision and gap-filling synthesis and ligation. In total there are ∼25−30 distinct polypeptides that function in GG-NER; seven of them are encoded by genes (*XPA* through *XPG*) in the xeroderma pigmentosum gene family (*12*).

In addition to its critical role in DNA damage sensing, XPC performs protein recruitment functions essential for effective GG-NER. This includes the recruitment of transcription factor IIH (TFIIH) via direct protein−protein interactions (*5*, *13*). The helicase activity of TFIIH subunits XPB and XPD prepares the lesion site for a transition from DNA damage sensing to DNA damage excision. The NER factor XPA is also believed to play a critical role in this transition, in association with RPA, by helping to dissociate the XPC−RAD23B complex from the damaged DNA (*14*)

† This work was supported by grants from the National Institutes of Health (RO1 GM65484 and RO1 GM40120 to W.J.C.; RO1 GM52948 to E.F.; P50 ES00267 to the Vanderbilt Center in Molecular Toxicology and P30 CA 68485 to the Vanderbilt Ingram Cancer Center for support of facilities; T32 GM07347 to the Vanderbilt Medical Scientist Training Program to C.G.B.) and from the Howard Hughes Medical Institute Professors Program (52003905 to E.F.).

* Address correspondence to this author. Tel: 615-936-2210. Fax: 615-936-2211. E-mail: walter.chazin@vanderbilt.edu.

‡ Departments of Biochemistry and Chemistry and Center for Structural Biology, Vanderbilt University.

§ Department of Biological Sciences, Vanderbilt University.

[1] Abbreviations: GG-NER, global genome nucleotide excision repair; UV, ultraviolet; 6-4PP, 6−4 photoproduct; CPD, cyclobutane pyrimidine dimer; XPC, xeroderma pigmentosum complementation group C; XPC-N, N-terminal portion of XPC [XPC(1−491)]; XPC-C, C-terminal portion of XPC [XPC(492−940)]; TC-NER, transcription-coupled nucleotide excision repair; XP, xeroderma pigmentosum; ERCC1, excision repair cross-complementing 1; RAD23B, human homologue of Rad23 B; TFIIH, transcription factor IIH; PNGases, peptide-*N*-glycanases; UV-DDB, UV-DNA damage binding protein complex; NFBA, nitrocellulose DNA filter binding assay; CETN2, human centrin-2; NTD, N-terminal domain; CTD, C-terminal domain; NMR, nuclear magnetic resonance spectroscopy; HSQC, heteronuclear single-quantum coherence; MSA, multiple sequence alignment; IPTG, isopropyl thio-β-D-galactopyranoside; PMSF, phenylmethanesulfonyl fluoride; BME, 2-mercaptoethanol; GuHCl, guanidine hydrochloride; MWCO, molecular weight cutoff; SUMO, small ubiquitin modifier; MBP, maltose binding protein; His_6, six-histidine N-terminal tag; CV, column volumes; PICT, protease inhibitor cocktail tablet.

10.1021/bi061370o CCC: $33.50
Published on Web 11/24/2006

and positioning the 3′- and 5′-nucleases XPG and XPF-ERCC1, respectively, for their damage excision activities (*15*−*17*).

Patients with the autosomal recessive disease xeroderma pigmentosum (XP) suffer decreased DNA repair capability because of abnormalities in one of the XP genes and its translated protein (*2*). XP patients are classified into specific complementation groups (XP-A through XP-G) depending on which XP gene is abnormal. Patients with XP experience 1000-fold higher incidences of basal and squamous cell carcinomas and malignant melanomas of the skin and more rarely, especially for XP-C patients, neurological symptoms (*18*−*20*). Different sensitivities to UV radiation and different susceptibilities to tumor development are observed among the XP complementation groups (*21*). Hence, the molecular basis for XP disease appears to be different for the different XP proteins; little is known about how these defects are related.

The importance of XPC is underscored by known mutations in the protein that are directly associated with XP disease (*22*). However, until the study reported here, there was no direct evidence for the biochemical defect(s) associated with any specific mutation found in XP-C patients. The medical relevance of XPC extends beyond just XP disease, as recent studies implicate the *XPC* gene in lung tumorigenesis (*23*).

Significant insights have been made recently into the biological integration of GG-NER and cell cycle processes. This includes not only more detailed understanding of the biochemical steps in GG-NER (*16*, *24*, *25*) but also the potential coordination of these steps with other cellular activities such as centrosomal cycling (*11*, *26*−*28*), proteasomal degradation (*29*, *30*), chromatin remodeling (*31*, *32*), and posttranslational modification such as sumoylation or ubiquitylation (*33*−*36*). XPC may have a central role in this network, since it interacts with multiple DNA processing factors: DNA (*37*), RAD23B (*37*), CETN2 (*26*), TFIIH (*5*), UV-DDB (*33*), XPA (*14*), and thymine DNA-glycosylase (*38*). The interactions with DNA, RAD23B, CETN2, and TFIIH have all been mapped to specific regions in the C-terminal half of XPC (residues 492−940) (*6*, *27*), but the interaction sites for the other binding partners are not known. Remarkably, despite being nearly 500 residues in length, there are no specific functional roles or structural domains assigned to the N-terminal portion of XPC (residues 1−491).

Although there is widespread interest in XPC, biochemical and structural characterization of this protein has been rather limited because it is large and very difficult to produce in high quantities by traditional methods. We report here the construction of high level bacterial expression systems for XPC fragments and their use to (1) identify a structural domain in the N-terminal half of XPC, (2) show that this domain interacts with XPA, (3) produce a stable construct containing the C-terminal portion of the protein (XPC-C), (4) show that XPC-C replicates the DNA-binding activity and certain protein interaction functions (RAD23B and CETN2) of the full-length protein, (5) determine the biochemical defect associated with a mutation found in certain patients with XP disease, and (6) construct and interpret a model of a folded XPC core.

EXPERIMENTAL PROCEDURES

Sequence Analysis. Five XPC homologues were selected for multiple sequence alignment (MSA) from an initial BLAST search performed at the Swiss Institute of Bioinformatics using the BLAST network service (*39*, *40*). Selection criteria for inclusion in the MSA were score and *e*-value and that the sequence was not flagged as a fragment, hypothetical, or predicted. The primary sequences of the XPC homologues [*Homo sapiens*, *Mus musculus*, *Drosophila melanogaster*, *Schizosaccharomyces pombe*, and *Saccharomyces cerevisiae*] were aligned using T-COFFEE (*41*) at the ExPASy proteomics server (*40*).

Two other prediction algorithms were accessed through the ExPASy proteomics server: (1) SSpro (*42*) was used for secondary structure prediction and (2) NetPhos 2.0 (*43*) was used for predictions of serine, threonine, and tyrosine phosphorylation sites, both using the human XPC primary sequence. Protein disorder predictions were performed using two algorithms that resulted in very similar predictions, DISOPRED (*44*, *45*) and IUPred (*46*).

Molecular Biology of Plasmid Constructs. Human XPC and CETN2 constructs were produced in three standard steps: (1) the desired DNA was synthesized by polymerase chain reaction (PCR) and purified by agarose gel electrophoresis, (2) the DNA was digested with *Nde*I and *Bam*HI (for the pET15b vector) or *Bam*HI and *Not*I (for the pLM302 and pBG102 vectors), and (3) the digested DNAs were subcloned into the appropriate restriction sites of either the pET15b (His_6), pLM302 (His_6-MBP), or pBG102 (His_6-SUMO) vectors. Both pLM302 and pBG102 are in-house modified vectors derived from pET27 (L. Mizoue, Center for Structural Biology, Vanderbilt University). The mutation Trp690Ser was incorporated into XPC-C using the QuikChange site-directed mutagenesis kit (Strategene, La Jolla, CA). In order to remove the single natural *Bam*HI restriction site in XPC (to allow for ligation into pET15b) nucleotides 1023 (G to A) and 1026 (T to C) were mutated, which resulted in no alteration of the sequence.

XPA constructs were produced by digesting the PCR DNA product with *Bam*HI and *Not*I and subcloning into either pGEX-6p-1 (GST), pLM302, or pBG102 vectors (all tags N-terminal). The RAD23B plasmid used in this study was kindly provided by Prof. Rick Wood (University of Pittsburgh Cancer Institute, Pittsburgh, PA) and fuses a C-terminal His_6 tag to RAD23B (*47*). DNA sequencing verified that all plasmids contained the correct nucleotide sequence.

Protein Expression and Isolation. XPC [Rosetta(DE3)], XPA [Rosetta(DE3)], RAD23B [BL21(DE3)], and CETN2 [BL21(DE3)] proteins were expressed in the indicated *Escherichia coli* strain at 37 °C. Typically, protein expression was induced with 1 mM isopropyl thio-β-D-galactopyranoside (IPTG) and proceeded for 2−4 h at 37 °C or for 4−6 h at 18−24 °C. Cells were harvested by centrifugation at 8000*g* and 4 °C. Cells were lysed by sonication at 4 °C in 20 mM Tris buffer (pH 8.0) containing 300 mM NaCl, 20 mM imidazole, 1% Nonidet P-40 (NP-40), 0.1 mg/mL lysozyme, ∼10 mg of DNase I, 5 mM 2-mercaptoethanol (BME), 0.5 mM phenylmethanesulfonyl fluoride (PMSF), and 10% protease inhibitor cocktail tablet (PICT; Roche Diagnostics, Indianapolis, IN) per liter of cell culture. The lysate was separated into soluble (XPC-N, RAD23B, CETN2,

XPA) and insoluble (XPC-C) fractions by centrifugation at ~23000*g* and 4 °C. For XPC(154−331) the pH was adjusted to 7.0 to enhance stability and solubility. For XPA proteins, the solution contained additional 20 μM zinc acetate and 1.0 mM magnesium sulfate with buffer at pH 7.0. Inclusion bodies of the XPC-C fragments were stored at −30 °C for refolding.

Refolding of XPC-C. Prior to refolding the XPC C-terminal fragments, the inclusion bodies were washed for 1 h at room temperature using 30 mL of detergent solution containing 20 mM potassium phosphate (K_2HPO_4) (pH 8.0), 100 mM potassium sulfate (K_2SO_4), 2% NP-40, 2% Triton X-100, 2% glycerol, 10 mM BME, and 0.5 mM PMSF. Insoluble material was pelleted by centrifugation for 10 min at ~23000*g* and 4 °C. The insoluble fraction was suspended with 30 mL of 20 mM Tris buffer (pH 7.0) containing 1.0 M guanidine hydrochloride (GuHCl), 300 mM NaCl, 20 mM imidazole, 10% glycerol, 1% NP-40, 5 mM BME, and 0.5 mM PMSF, and the remaining insoluble material was pelleted by centrifugation for 10 min at ~23000*g* and 20 °C.

The protein in the pellet was denatured in 30 mL of 20 mM Tris buffer (pH 7.0) containing 6 M GuHCl, 300 mM NaCl, 20 mM imidazole, 10% glycerol, 1% NP-40, and 5 mM BME [solid-phase refolding solution 1 (SPR1)]. The solution was applied to a 6−8 mL Ni^{2+} affinity column, washed with ~15 CV of SPR1 (room temperature) followed by another ~45 CV of solid-phase refolding solution 2 (SPR2) (4 °C), which contained 20 mM Tris buffer (pH 7.0), 300 mM NaCl, 20 mM imidazole, 10% glycerol, 1% NP-40, 5 mM BME, 0.5 mM PMSF, and 1 PICT (per 250 mL). The protein was eluted from the Ni^{2+} resin using SPR2 containing increasingly higher amounts of imidazole (40, 80, 120, 160, 200, or 500 mM) and sufficient solid arginine in the tube to yield a 200 mM concentration in the final eluate volume. The first 500 mM imidazole elution contained the majority of the refolded XPC-C (~20 mg final yield).

Protein Purification. Both RAD23B and CETN2 constructs were purified by Ni^{2+} affinity chromatography followed by Q−Sepharose ion-exchange chromatography (5−10 mg final yield). His_6-tagged N-terminal XPC proteins were first purified by binding to HIS-Select nickel affinity resin (Sigma, St. Louis, MO) and eluting with increasing imidazole gradient. A second purification step included either anion- or cation-exchange chromatography using DEAE- or SP-Sepharose HiTrap columns (Amersham Biosciences, Uppsala, Sweden), respectively. GST−XPA was purified by binding to GSH−Sepharose resin (Amersham Biosciences).

Limited Proteolysis and Protease Protection. Limited proteolysis was performed at room temperature over 2 h using 130 μL of purified XPC(65−331) or XPC(492−940) solution from the primary refolding elution (~0.4 mg/mL). Either proteinase K, subtilisin, chymotrypsin, or trypsin (3−20 ng) (from 1 ng/μL stock solutions) was added to the samples and incubated. Aliquots of 10 μL were removed from the reaction mix at time points 0, 1, 5, 10, 15, 20, 30, 45, 60, 90, and 120 min postaddition of the protease and combined with 10 μL of gel loading solution. The gel sample for each time point was heated immediately at 90 °C for 10 min and then placed on ice for the remainder of the experiment until complete analysis by SDS−PAGE.

Protease protection assays were performed in a similar manner as the limited proteolysis reactions. XPC(65−331) was supplemented with either His-SUMO-XPA or mock buffer [20 mM Tris (pH 7.5), 300 mM NaCl, 200 mM imidazole, 20 μM zinc acetate, 1.0 mM magnesium sulfate, 1 mM BME]. XPA was added to give a 1:1 stoichiometry with XPC(65−331). Control XPA digestions were also performed. Chymotrypsin (5 ng) was added to each sample and the assay carried out as detailed for limited proteolysis. Additional time points at 24, 48, 72, and 96 h after protease addition were analyzed to monitor the long-term stability conferred by the interaction of the proteins.

XPC(492−940) was supplemented with either CETN2 or just buffer. CETN2 was added in an ~1:1 stoichiometry with XPC-C (both proteins at ~7.5 μM). A control CETN2 digestion was also performed. Proteinase K (5.5 ng) was added to each sample and the assay carried out as detailed for limited proteolysis. An additional time point at 18 h after protease addition was analyzed to monitor the long-term stability conferred by XPC-C−CETN2 interaction. Protease protection experiments between XPC-C and RAD23B were performed in an identical manner.

ssDNA−Cellulose Binding. Refolded XPC(492−940) was first dialyzed at 4 °C into 1 L of a 25 mM Tris buffer (pH 7.0) containing 1 mM EDTA, 10% glycerol, 0.01% NP-40, 5 mM BME, 0.5 mM PMSF, and NaCl (2.0−0.5 M by 0.5 M steps). RAD23B and CETN2(22−172) were prepared by dialysis with XPC(492−940) during the final dialysis step. The final concentrations of the XPC(492−940), RAD23B, and CETN2(22−172) stock solutions were determined by a combination of UV-absorption spectrophotometry and amino acid analysis. Protein samples were prepared the day before the experiment, as detailed in the legend to Figure 7.

Cellulose bound to calf thymus ssDNA (Sigma, St. Louis, MO) was prepared according to the methods described by Alberts and Herrick (*48*). Columns were equilibrated into one of two binding solutions: (1) for experiments in Figure 7, 20 mM Tris, 30 mM HEPES (pH 7.0), 500 mM NaCl, 7 mM $MgCl_2$, 100 μM $ZnCl_2$, 5 mM creatine phosphate, 2 mM ATP, 2.7% glycerol, 0.0027% NP-40, 5 mM BME, and 0.5 mM PMSF; (2) for experiments in Figure 8, 20 mM Tris (pH 7.0), 500 mM NaCl, 5 mM BME, and 0.5 mM PMSF.

Protein samples were applied to either the ssDNA−cellulose or cellulose columns by gravity flow at 4 °C. The columns and any bound protein were washed, and bound proteins were eluted from the columns using binding solution containing 2.0 M NaCl rather than 0.5 M NaCl.

Nitrocellulose−DNA Filter Binding Assay. Nitrocellulose disks (0.45 μm) (Millipore, Billerica, MA) were prepared for protein retention by base treatment with 0.5 M potassium hydroxide solution for 20 min (*49*). The oligomer (MWG-BIOTECH Inc., High Point, NC) (5′-TTCCTCCTTCT-TCTCTTCCTCCTTCTTGGTTCTCTTCTTCTC CTCTTC-CTCCTTCTTCTC-3′) was radiolabeled using polynucleotide kinase. UV-damaged DNA was prepared by irradiating ^{32}P-DNA with 20 kJ/m^2 using a Stratagene Stratalinker (Stratagene, La Jolla, CA).

Binding reaction samples were prepared by mixing protein (0−400 nM), DNA (30−40 nM), and buffer to achieve a final reaction volume of 15 μL and final solution composition of 40 mM creatine phosphate, 30 mM HEPES (pH 7.9), 7 mM $MgCl_2$, 4 mM Tris, 2.7% glycerol, 0.0027% NP-40, 100

mM NaCl, and 1.0 mM BME (for Figure 7E) or 23 mM Tris (pH 7.0), 0.09 mM EDTA, 9.3% glycerol, 0.009% NP-40, 460 mM NaCl, and 4.6 mM BME (for Figure 8E). The samples were then centrifuged at 14000*g* for 5 s, incubated at room temperature for 30 min, and then placed on ice prior to application to the disks. Treated nitrocellulose disks were placed on a vacuum apparatus (Hoefer Scientific Instruments Model FH225V, San Francisco, CA; kindly provided by Prof. Wallace LeStourgeon, Department of of Biological Sciences, Vanderbilt University), sealed using 25 mm diameter chimney weights, and prewashed with 1.0 mL of a solution containing 30 mM HEPES (pH 7.9) and 7 mM $MgCl_2$ (NFBA solution). A single binding reaction sample then was applied to a disk under vacuum (20−30 mmHg). After application of the sample, each disk was washed with 10 mL (5 × 2 mL) of NFBA solution. Upon completing the wash steps, the chimney weights and disks were removed and the disks allowed to air-dry for 1 h before being immersed in 3 mL of Betamax ES scintillation fluid (MP Biomedicals, Irvine, CA). Scintillation counts were obtained using a Beckman LS5000TD (Fullerton, CA) apparatus.

Saturation isotherm analysis was performed by fitting the binding curves to a simple two-state binding equation: $m_3 + [(m_1 m_0)/(m_2 + m_0)]$, where m_0 is the total protein concentration, m_1 is the curve amplitude, m_2 is the dissociation constant, and m_3 is the *y*-value (bound protein) at [protein] = 0. Curve fitting analysis and the plotting of figures were performed using Kaleidagraph 3.5 (Synergy Software, Reading, PA).

Circular Dichroism. Circular dichroism (CD) measurements were made using samples that were 30−50 μM protein in either 20 mM Tris buffer (pH 8.0) containing 300 mM NaCl, 1 mM BME, 0.5 mM PMSF [XPC(65−331), and XPC(65−128)] or 100 mM sodium phosphate buffer (pH 7.0) containing 150 mM NaCl and 1 mM DTT [XPC(154−331)]. A Jasco J-810 CD spectrometer (Easton, MD) was used to scan samples in either a 0.1 cm (far-UV CD) or 1 cm (near-UV CD) path length cuvette from wavelength 250 to 200 nm (far-UV) or 350 to 240 nm (near-UV). Temperature was controlled at 4 °C using a Peltier device.

GST Pull Down. XPC fragments [XPC(65−331), XPC-(344−517), MBP-XPC(492−940), XPC(815−940)] were expressed in 10 mL LB cultures, and cell pellets were suspended in 20 mM Tris buffer (pH 7.0) containing 300 mM NaCl, 20 mM imidazole, 1 mM BME, 1% NP-40, 20 μM zinc acetate, 1.0 mM magnesium sulfate, 0.5 mM PMSF, 0.1 mg/mL lysozyme, ∼10 mg of DNase I, and 0.1% PICT. Cell suspensions were lysed by sonication at 4 °C. A 1 mL aliquot of each sonicated protein solution was transferred to an Eppendorf tube and separated into soluble and insoluble fractions by microcentrifugation for 5 min at 16100*g* and 4 °C. The soluble fractions were carefully pipetted into fresh Eppendorf tubes containing 35 μL of GSH−Sepharose resin bound to GST−XPA. The XPC/GSH−Sepharose−GST−XPA mixes were gently rocked for 1 h at 4 °C and then microcentrifuged for 1 min at 1000 rpm and 4 °C. Supernatants were decanted, and the GSH−Sepharose−GST−XPA resin was treated with five cycles of washing with 1 mL of binding solution [20 mM Tris buffer (pH 7.0) containing 300 mM NaCl, 20 mM imidazole, 1 mM BME, 1% NP-40, 20 μM zinc acetate, 1.0 mM magnesium sulfate, 100 μg/mL bovine serum albumin, and 0.5 mM PMSF] followed by microcentrifugation for 1 min at 1000 rpm and 4 °C. After washing, the supernatant was carefully removed by pipet, and proteins bound to the GSH−Sepharose−GST−XPA resin were analyzed by SDS−PAGE. Control experiments were performed using only GSH−Sepharose−GST resin. Additional GSH−Sepharose−GST−XPA pull-down experiments were performed using highly purified XPC(65−331), XPC(65−128), and XPC(154−331) rather than cellular lysate.

Modeling of XPC-TGC. The remnant N- and C-terminal transglutaminase sequences were fused in silico, omitting ∼150 residues of the intervening linker, to create a 319-residue XPC transglutaminase core (XPC-TGC) sequence that fused residues 154−331 to residues 517−657. Twenty-five residues from the linker between the remnant transglutaminase sequences were left in place in order to obtain insight into the probable spatial location of the long unfolded segment relative to the N- and C-terminal subdomains. The fusion sequence was analyzed by the Meta Server at the Bio Info Bank Institute (www.bioinfo.pl) to identify the most suitable structural template and generate a sequence alignment to guide the modeling. The yeast PNGase X-ray crystal structure (*50*) scored as the best template for modeling XPC-TGC (*J* score = 142). The sequence alignment was used to generate a homology model of XPC-TGC using MODELLER (*51*) on the Bioinformatics Toolkit Server (Max Planck Institute) (*52*). Residues contained in the final model include 168−331 and 517−646 with 331 fused to 517, resulting in a 294 amino acid XPC-TGC construct. The side chains of the model were repacked, and the model was energy minimized using SCWRL (version 3.0) (*53*), followed by additional energy minimization and molecular dynamics calculations using AMBER (version 7.0) (*54*). PROCHECK analysis (*55*) indicates that the model has stereochemical parameters as good as the template crystal structure with 99.2% of the backbone dihedral angles in the allowed regions of ϕ,ψ space. Further analysis of the packing of residues in the model was performed using QPACK (sumpairs score = −1.7) (*56*), which indicates a high-quality packing arrangement. Superposition of XPC-TGC and the yeast PNGase crystal structure gives a backbone rms deviation of 1.1 Å.

RESULTS

The lack of availability of full-length XPC protein has greatly hindered efforts to characterize its structure and function. XPC does not express in *E. coli* (*57*), and fragments of XPC that do express are mostly found in insoluble inclusion bodies (*58*). Several reasons have been suggested for these bacterial production difficulties: (1) high concentrations of XPC may be toxic to *E. coli*, possibly through altered DNA metabolism (*59*, *60*); (2) XPC possesses a high number of arginine codons whose tRNAs are rare in *E. coli* (*10*); (3) XPC is intrinsically unstable without binding partners and thus susceptible to proteolytic degradation (*29*, *57*). Hence, the first goal was to overcome these challenges and develop reagents for subsequent biochemical and structural analysis.

Initial efforts to characterize the structure of XPC focused on producing full-length protein for limited proteolysis experiments combined with N-terminal sequencing and mass spectrometry. Despite extensive efforts to use baculovirus

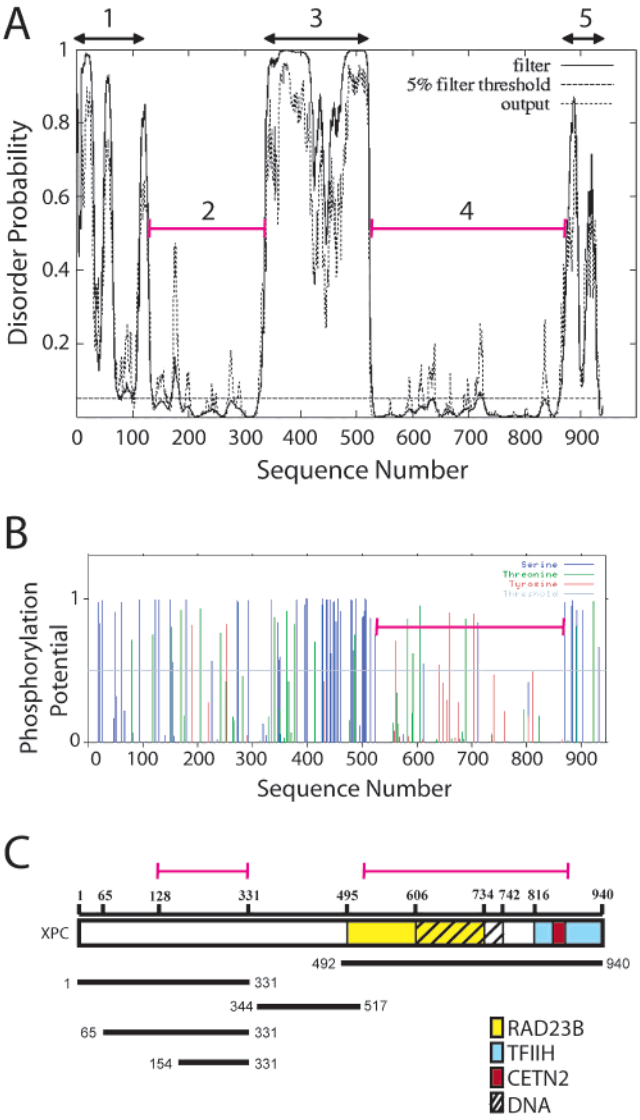

FIGURE 1: Analyses of the human XPC amino acid sequence. (A) Plot of protein disorder probability versus residue. The red bars indicate regions of XPC predicted to be highly ordered/structured. (B) Plot of phosphorylation potential versus residue for serine (blue), threonine (green), and tyrosine (red) residues. The red bar indicates a region of XPC that is distinct because of its lack of high potential serine phosphorylation sites. (C) Schematic diagram of XPC domains and several constructs designed from sequence analyses. The XPC regions proposed to interact with RAD23B (*6*), TFIIH (*5*), CETN2 (*27*), and DNA (*6*) are highlighted as indicated. Red bars indicate the predicted ordered regions from panel A.

expression on a larger scale than previously described (*10*), yields of soluble XPC from baculovirus infection of Sf9 insect cells were so low that a sufficient amount of full-length protein could not be generated. We therefore turned to an alternative strategy, combining sequence analysis with a series of subcloning and biochemical experiments.

Analysis of the XPC Sequence. Four different algorithms were used to analyze the sequence of XPC: (1) protein disorder prediction; (2) phosphorylation potential prediction; (3) secondary structure prediction; (4) multiple sequence alignment (MSA) using five homologues (*H. sapiens*, *M. musculus*, *D. melanogaster*, *S. pombe*, and *S. cerevisiae*). Application of these computational tools suggested the presence of at least one structured region in XPC-N and one or more additional structured regions in XPC-C (Figure 1).

Protein disorder prediction suggests an ordered region exists in XPC-N (segment 2), which is flanked by disordered regions (segments 1 and 3) (Figure 1A). Secondary structure predictions indicated that segments 1 and 2 both contain approximately 50% α-helix, 10% β-strand, and 40% coil, whereas segment 3 is predicted to contain 9% α-helix, 3% β-strand, and 88% coil. MSA did not reveal any large regions of primary sequence conservation spanning the N-terminal portion of the protein. However, two short regions within segment 2 do have significant similarity to the homologous proteins (vide infra).

MSA proved to be the most powerful approach for analyzing XPC-C because a large conserved region was found spanning residues 530−870 (~15% sequence identity and 41% sequence similarity; Supporting Information, Figure 1). Secondary structure prediction for this conserved region gives 39% α-helix, 16% β-strand, and 45% coil. Prediction of protein disorder and phosphorylation site potential reinforced the MSA results: disorder probability was near zero between residues 520 and 870 (Figure 1A, segment 4), and the phosphorylation potential distribution over that same amino acid span was remarkably different from the rest of the protein (Figure 1B). In particular, many serine residues in XPC-N showed strong phosphorylation potential (near 1.0), including a cluster of serines in the middle of the sequence, whereas the region between residues 524 and 869 was largely devoid of such serine sites. Notably, our results from sequence analysis were in agreement with the general predictions deposited in the Pfam-A database (*61*, *62*), which also suggest that the region 520−870 is fundamentally distinct from the rest of XPC. The prediction of structure in the region 520−870 is also fully consistent with the mapping of XPC interactions with DNA (*6*), RAD23B (*6*), and CETN2 (*11*, *27*) to functional domains within XPC-C (Figure 1C).

Protein Production and Analysis of Domains. Working from the sequence analysis, a number of constructs were designed around the putative structured regions (Figure 1A, segments 2 and 4). The primary goal in these designs was to produce domains that could be readily produced to map and characterize structure and function. Over 60 XPC constructs were expressed and purified; a complete list is provided in Supporting Information, Figure 2. Of all of the XPC-N constructs, XPC(65−331) was most promising as it demonstrated excellent expression and solubility properties. XPC-C constructs expressed primarily in inclusion bodies, so it was necessary to develop a protein refolding procedure (see Experimental Procedures; Supporting Information, Figure 3). The refolding protocol took advantage of the His_6 tag, which enabled use of an in situ refolding strategy (*63*). Of all of the XPC-C constructs, the full-length XPC(492−940) displayed the most efficient refolding and highest yield of soluble protein. Although many other constructs were characterized, the focus from this point on will be on these two "best" constructs: XPC(65−331) and XPC(492−940).

To determine if XPC(65−331) contained one or more structural domains, it was analyzed by a series of limited proteolysis experiments using four different proteases: trypsin, chymotrypsin, subtilisn, and proteinase K. Formation of an ~20 kDa fragment was observed in all experiments (Figure 2A, band b). To identify this fragment, it was gel excised and digested into peptides using excess trypsin or chymot-

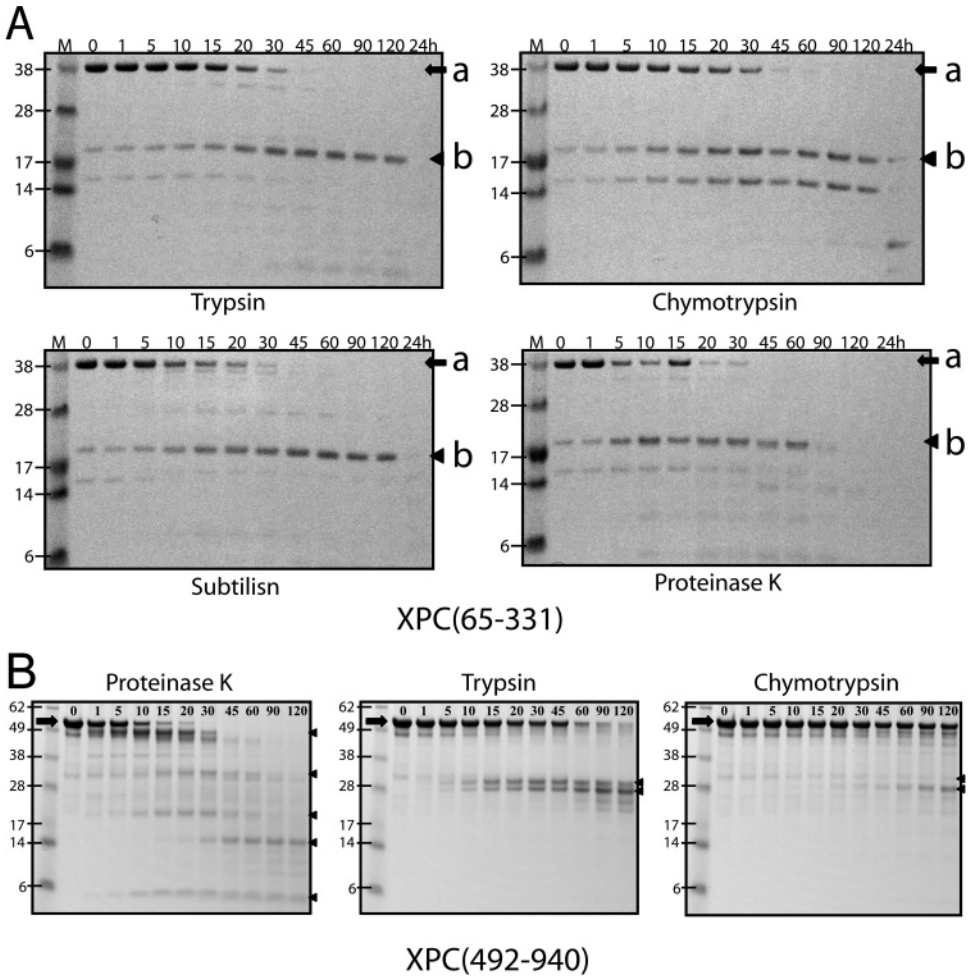

FIGURE 2: Characterization of the stability of XPC fragments by limited proteolysis. (A) XPC(65−331) proteolysis reaction monitored over the course of 2 h with an additional time point at 24 h. Band a represents the parent XPC(65−331) protein (black arrow), and band b represents the major ~20 kDa stable fragment found in all proteolysis reactions (arrowhead). (B) Investigation of the stability of XPC-(492−940). Arrows represent the intact protein, while arrowheads indicate the major proteolysis fragments observed during the experiment. 7.5 μL of each reaction was analyzed by SDS−PAGE (4−12%) and visualized for protein by staining with Simply Blue SafeStain (Invitrogen). Protein markers (M) are indicated in kDa.

rypsin, and the peptides were analyzed by MALDI-TOF MS/MS. The data showed that this fragment contains residues 156−325 (Supporting Information, Figure 4). An XPC(154−331) construct was then designed and produced for further analysis.

The presence of stable structure within XPC(65−331) was further assessed by circular dichroism (CD). Analysis of the far-UV CD spectrum for XPC(65−331) reveals primarily α-helical secondary structure, as evidenced by the characteristic minima at 208 and 222 nm (Figure 3A). The near-UV CD spectrum (Figure 3B) indicates a stable tertiary fold with significant aromatic interactions in the hydrophobic core. XPC(154−331) produced a CD spectrum with strong α-helical character (Figure 3C). In stark contrast, isolated XPC(65−128) produced a CD spectrum predominantly of random coil nature (Figure 3D). Thus, XPC(65−331) contains a structured domain spanning residues 154−331. These data are the first direct evidence for a folded structural domain within the N-terminal portion of XPC.

To assess structural stability and determine if there are distinct structural domains in the C-terminal portion of XPC, limited proteolysis experiments were performed on XPC-(492−940) using three different proteases: proteinase K, trypsin, and chymotrypsin (Figure 2B). Several discrete and stable fragments formed during incubation of XPC(492−940) with the nonspecific protease proteinase K (177 potential cleavage sites), suggesting the presence of structured regions (Figure 2B, left panel). In the case of the more specific proteases trypsin and chymotrypsin (71 and 86 potential cleavage sites, respectively), XPC(492−940) exhibited a remarkable resistance to proteolysis over the duration of the experiments (Figure 2B, middle and right panels). Figure 2B shows that, in addition to the persistence of intact XPC(492−940), two major fragments in the range 25000−30000 Da formed during trypsin and chymotrypsin digestion. These observations suggest that the proteases were only able to cleave at a select number of exposed residues, while other sites were buried, consistent with the presence of a stable folded core within XPC-C.

Structural Model of the XPC Core. The evidence for the presence of stable folded structural domains in both the N- and C-terminal portions of XPC prompted further investigation of the sequence, focusing explicitly on the putative structured regions. As noted above, XPC contains two regions (154−331, 517−657) that align with members of the transglutaminase superfamily (transglutaminases, PNGases, and XPC/Rad4) (*64*). Remarkably, the remnant transglutaminase sequence in XPC is divided into two parts

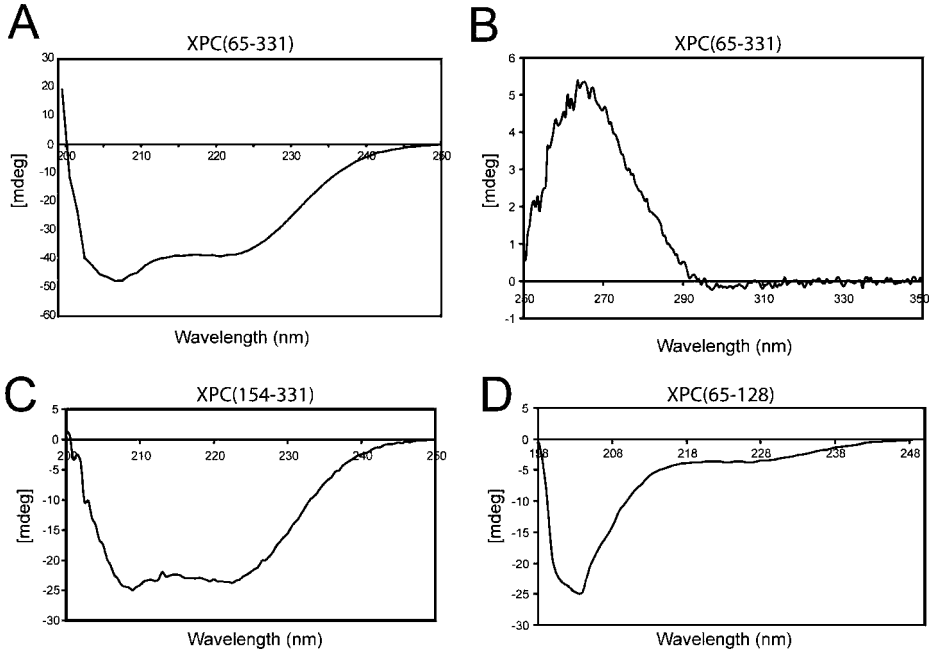

FIGURE 3: Identification of a structured N-terminal domain in XPC using circular dichroism. (A, B) The far-UV (A) and near-UV (B) CD spectra of XPC(65−331), respectively. (C, D) The far-UV CD spectra of (C) XPC(154−331) and (D) XPC(65−128), respectively.

by an ~175 residue insertion, which corresponds closely to the predicted highly disordered segment 3 (Figure 1A). The 154−331 sequence corresponds closely to the N-terminal structural domain we discovered from biochemical analysis. The remainder of the transglutaminase sequence (residues 519−634) corresponds to the first third of the predicted structured region (segment 4) within XPC-C.

To gain further insight into the putative transglutaminase fold in XPC, a homology model of the XPC transglutaminase core (XPC-TGC) was constructed using the recently determined X-ray crystal structure of the yeast PNGase (*50*). To facilitate the model building, the N- and C-terminal transglutaminase sequences were fused in silico, reducing the intervening linker from ~175 to 25 residues (Figure 4A). Thus, a 319-residue construct fusing residues 154−331 to 517−657 was built and analyzed using the Meta Server at the Bio Info Bank Institute (*65*). The Meta Server sequence alignment (Supporting Information, Figure 5) was used to generate the homology model. The best template obtained was the yeast PNGase (J score = 142). Surprisingly, analysis of alignments from 154 to 331 or 517 to 657 alone achieved lower alignment scores than that for the fused sequence. Residues contained in the final model include 168−331 fused to 517−646, leaving out 26 residues from the alignment at the N- and C-termini, because information was not sufficient to model their structure. The quality of the model was assessed by a number of standard approaches as described in Experimental Procedures. For example, PROCHECK analysis revealed that the stereochemical parameters of the model are as good as the template, with 99.2% of the backbone dihedral angles in the allowed regions of ϕ,ψ space.

As expected, the XPC-TGC closely resembles the transglutaminase fold present in the yeast PNGase. The rmsd over all backbone atoms in the model is only 1.1 Å. The structure contains a highly α-helical N-terminal subdomain that packs with the β-strands of the C-terminal subdomain to generate a complete transglutaminase fold (Figure 4A). The strongest linkage between the two parts of the fold is the insertion of a β-strand from the N-terminal subdomain (Thr313−Ser318) into the C-terminal subdomain to form a four-stranded β-sheet in the manner of a domain-swapped dimer (Figure 4A). The interface between the two subdomains is substantial with ~3200 A^2 of occluded accessible surface area. The 25-residue linker forms a loop projecting out from the surface of XPC-TGC, implying that the complete 175-residue linker in native XPC could also extend from the surface of the protein without interfering with the (putative) folded core.

The molecular surface of the XPC-TGC was additionally analyzed for its hydrophobic character and electrostatic potential (Figure 4B). Overall, the molecular surface of XPC-TGC is unremarkable for any distinct large hydrophobic patches; however, there are unique electrostatic surface features. For example, the central β-sheet and an α-helix (Asp295−Leu309) form the ends of an acidic crevice (Figure 4B). The acidic crevice of the yeast PNGase not only contains the catalytic triad residues necessary for its glycanase activities but even had bound sucrose molecules in the crystal structure (*50*). Since XPC does not have a functional catalytic triad, the role of the acidic crevice is unknown. Certainly, the acidic nature of the crevice would not be conducive to binding to DNA because of repulsions with the negatively charged phosphodiester backbone. Thus, the acidic crevice presumably serves to mediate protein−protein interactions. The molecular surface of XPC-TGC is notable also for several basic patches (Figure 4B). The overall isoelectric point of XPC is 9.1, consistent with the large regions of positive charge on the surface, which point to binding of DNA. However, only residues 606−646 in the model have been implicated in DNA binding with the remainder extending beyond toward the C-terminus (*6*). Hence, the majority of the positively charged surface of XPC-TGC apparently serves functions other than DNA binding.

XPC(154−331) Interacts with XPA. As noted above, there are no confirmed functions currently assigned to the N-terminal portion of XPC. A survey of interactions of XPC-N

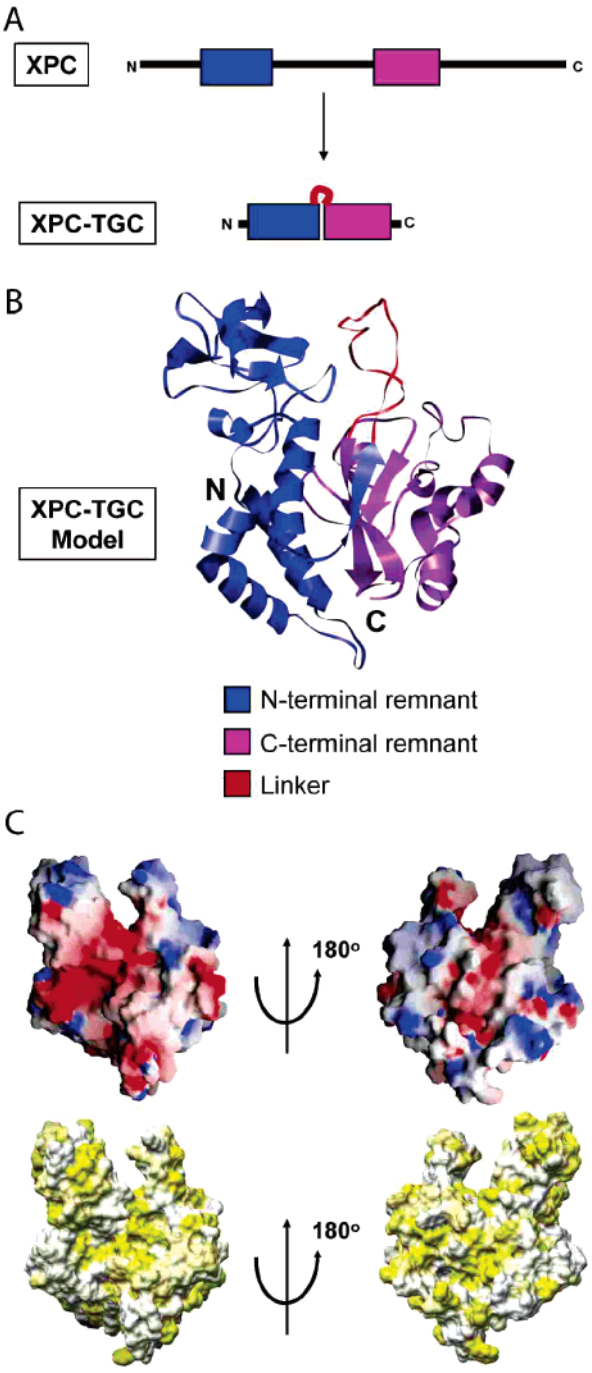

FIGURE 4: Structural model of the putative XPC transglutaminase core. (A) Schematic diagram of the XPC-TGC sequence. (B) Ribbon diagram of the homology model with the two subdomains and linker color coded. (C) Molecular surface representations of XPC-TGC colored according to electrostatic potential (top) and hydrophobic accessible surface (bottom). Negative charge potential is represented in red, positive charge potential in blue, and hydrophobic potential in yellow. The orientation of the XPC-TGC surface on the left side corresponds to the same orientation as shown in the ribbon diagram in panel B. Surfaces were prepared using GRASP (top) (*78*) and Chimera (bottom) (*79*).

with ssDNA, RAD23B, and PCNA (data not shown) did not demonstrate positive hits for binding. However, preliminary experiments did suggest that further investigation of possible interactions with XPA was warranted. XPA was previously determined to physically interact with XPC (*14*, *66*), but the precise region of XPC involved in the interaction was not elucidated.

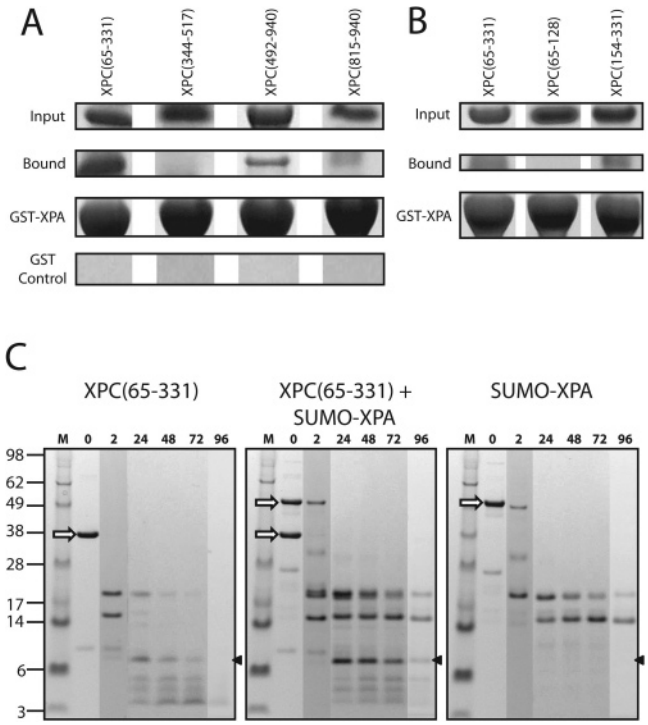

FIGURE 5: Characterization of the interaction of XPC(65−331) with XPA. (A, B) GSH−Sepharose resin with bound GST-XPA was used in pull-down experiments on the indicated XPC constructs. Input: 0.005% of the soluble fraction of the XPC construct was incubated with GSH−Sepharose−GST−XPA resin. Bound and GST control: 18% of the resin was analyzed for bound XPC. GST−XPA: 18% of the resin was used in the pull-down experiments. (C) Protease protection experiments performed with proteinase K. Time in hours is indicated at the top of each lane, and protein markers (kDa) are indicated on the left. White arrows denote the starting proteins. The arrowhead indicates the significantly protected XPC fragment.

To determine if it is specifically the structured domain XPC(154−331) that is required for interaction with XPA, a series of XPA pull-down experiments were performed with a variety of different XPC constructs. One set of experiments involved using GST−XPA bound to GSH−Sepharose resin to pull down overexpressed XPC fragments directly from the bacterial cellular lysate, i.e., without purification. Figure 5A shows that GST−XPA has a significant affinity for XPC-(65−331), a significantly weaker interaction with XPC(492−940), and no interaction with regions of XPC that are predicted to be disordered (segments 1, 3, and 5 in Figure 1A). A second set of pull-down experiments were then performed using highly purified XPC(65−331), XPC(65−128), and XPC(154−331). As observed directly from cellular lysate, XPA showed physical interaction with XPC(65−331) (Figure 5B). The region of interaction was furthered refined to residues 154−331, since XPA was able to pull down XPC-(154−331) but not XPC(65−128) (Figure 5B). One additional qualitative observation was made: comparison of our pull-down results to published data suggested that the overall affinity of the interaction of XPC with XPA is weaker or more transient than the high-affinity interactions of native XPC with RAD23B and CETN2 (1:1 stoichiometry) (*27*, *37*).

XPA interactions with the N-terminal region of XPC were further examined by protease protection experiments using XPC(65−331) and His_6-SUMO-XPA digestion with chymotrypsin. In the absence of a binding partner intact XPC-(65−331) was digested within 2 h into two prominent

fragments between 15 and 20 kDa. By 48 h only a few faint fragments <8 kDa were left (Figure 5C, left panel). When XPA alone was digested, two stable fragments also between 15 and 20 kDa were formed and remained stable over 96 h (Figure 5C, right panel). When XPC(65−331) and XPA were incubated together, there was a striking protection of an ∼8 kDa fragment (Figure 5C, middle panel). The effect was particularly noticeable at 96 h, where no fragment of this size was observed in either of the experiments with isolated proteins. To identify the origin of this fragment, the band was excised from the gel and subjected to MALDI-TOF MS/MS analysis, which showed that it contains residues 248−325. These data thus confirmed that the fragment protected by binding of XPA arises from the structured domain of XPC-N. To further confirm these findings, the protease protection experiments were repeated using a MBP fusion (His_6-MBP-XPA), and essentially identical results were obtained (Supporting Information, Figure 6). Together, the pull-down and protease protection experiments indicate that the interaction of XPC with XPA is primarily mediated by the structured N-terminal domain XPC(154−331).

XPC-C Physically Interacts with CETN2 and RAD23B. To investigate the biochemical functions of XPC(492−940), a series of protease protection, pull-down, ssDNA−cellulose affinity, and filter binding assays were performed to probe for protein−protein and protein−DNA interactions. Physical interaction with CETN2 and RAD23B is known to stabilize native XPC and stimulate its ability to bind DNA (*6*, *11*), and the interaction sites have been mapped to the C-terminal half of the protein. Protease protection experiments were performed with XPC(492−940) to test for physical interactions with CETN2 and RAD23B. Figure 6A shows the stability to digestion by proteinase K of CETN2 and XPC(492−940) alone and in complex. In these experiments, isolated CETN2 was degraded within 30 min by proteinase K (Figure 6A, right panel). The fragmentation of XPC(492−940) in the absence of CETN2 was identical to the proteolysis pattern observed in Figure 2B (Figure 6A, left panel). When CETN2 and XPC(492−940) were incubated together with proteinase K, a large fragment of CETN2 displayed a remarkable resistance to proteolysis, persisting for 18 h after the start of the digestion reaction (Figure 6A, middle panel). This fragment (∼17 kDa) is consistent with the proteolytic loss of the flexible 21 amino acid N-terminal extension of CETN2. The persistent fragment of CETN2 contains the N- and C-terminal globular domains, suggesting that both domains and/or the linker between them were protected against proteolysis by a region of XPC-C. The nearly complete digestion of XPC-(492−940) at 18 h despite the presence of CETN2 suggests that the CETN2 binding region of XPC is small in size. This is consistent with other studies mapping the CETN2 interaction site of XPC to a 17-residue region (847−863) (*11*, *27*). The formation of a complex between CETN2 and XPC(492−940) was further corroborated by NMR studies, which showed binding of CETN2 to several of the XPC-C constructs (Supporting Information, Figures 7−9).

A similar strategy was used to examine the interaction of RAD23B with XPC(492−940). When free RAD23B was digested with proteinase K, the protein was degraded by 10 min into a few species >10 kDa and several small fragments around 6 kDa (Figure 6B, right panel). Treatment of XPC-(492−940) resulted in degradation into several fragments by

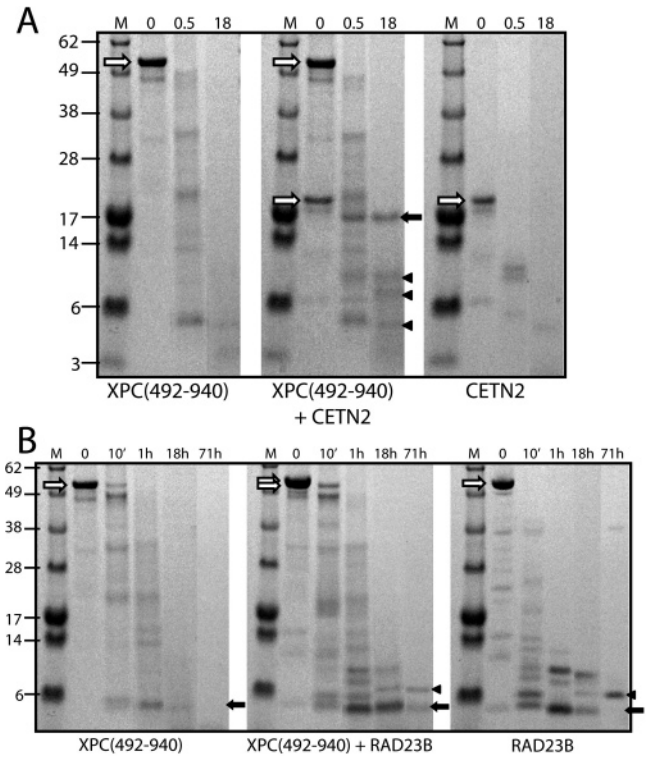

FIGURE 6: Characterization of XPC(492−940) interactions with CETN2 and RAD23B. (A) Stability of the complex with CETN2 monitored by protease protection. 5.5 ng of proteinase K was mixed with three different 190 μL reaction samples: 7.5 μM XPC-C (left), 7.5 μM CETN2 (right), and 7.5 μM XPC-C plus 7.5 μM CETN2 (middle). SDS−PAGE (10%) at 0, 0.5, and 18 h after the addition of proteinase K are shown, visualized using Simply Blue SafeStain. White arrows indicate the starting XPC-C or CETN2 protein, while the black arrow indicates the large CETN2 fragment. Arrowheads mark three other protected fragments arising from XPC or individual CETN2 domains. (B) Stability of the complex with RAD23B monitored by protease protection. The experiments were performed under conditions identical to those in panel A. The results at 0, 10 min, 1 h, 18 h, and 71 h after the addition of proteinase K are shown. White arrows indicate the starting XPC(492−940) or RAD23B protein. The black arrow indicates the major ∼5 kDa fragment stabilized by the presence of both proteins that is discussed in the text. The black arrowhead marks another highly stable fragment. Marker protein labels are in kDa.

10 min, although some intact protein was still present (Figure 6B, left panel). When XPC(492−940) and RAD23B were incubated together, the rate of degradation of the intact proteins was noticeably slowed (Figure 6B, middle panel). [Note that both intact proteins have the same initial size, hence the appearance of 2X total protein.] After 1 h of proteinase K digestion, both XPC(492−940) and RAD23B alone were degraded into small ∼5 kDa fragments, which was especially prominent for RAD23B (Figure 6B, left and right panels). The rate of degradation of these 5 kDa fragments was significantly slowed when XPC(492−940) and RAD23B were present together, as evidenced by the substantially larger bands at 18 h compared to the nearly absent bands at the same time for the isolated proteins. Moreover, the 5 kDa fragment is still present at 71 h for the complex, whereas there is no evidence of this fragment for either of the isolated proteins. Thus, as in the case of CETN2, protease protection suggests that XPC(492−940) physically interacts with RAD23B. Interestingly, previous work mapped the XPC-binding domain of RAD23B to a 56-residue region

(*47*), consistent with the 5−6 kDa fragments observed in these experiments (Figure 6B, middle panel).

XPC-C Binds ssDNA with High Affinity. XPC binds both ssDNA and dsDNA in a region mapped to the C-terminal portion of the protein (*4*, *37*, *57*, *67*−*69*). A ssDNA−cellulose affinity chromatography assay was first used to determine whether XPC(492−940) retained the DNA-binding functionality of the native protein. When XPC(492−940) was applied to a ssDNA−cellulose affinity column, it bound to the ssDNA−cellulose and was retained until elution with elevated NaCl concentration (2.0 M NaCl) (Figure 7A). The application of XPC(492−940) to a control column under identical solution conditions demonstrated that it specifically recognized the ssDNA component. Thus, XPC(492−940) possesses the ability to bind to ssDNA. The strength of the interaction was reflected in retention of XPC(492−940) on the column even with a starting concentration of 0.5 M NaCl.

To characterize the DNA-binding properties of XPC(492−940) in greater detail, nitrocellulose filter binding assays (NFBA) were performed. The DNA sequence was chosen so that the data could be compared to the previously reported measurements of the DNA-binding properties of native XPC (*14*). The results were sufficiently reproducible to enable binding constants to be determined. Indeed, titration of increasing amounts of XPC(492−940) to ^{32}P-60-nucleotide−ssDNA provided binding curves that could be fit using a simple two-state binding equation to estimate a dissociation constant (K_D). A value of 23 ± 6 nM was determined for undamaged DNA (Figure 7E). Titration of XPC(492−940) to UV-damaged ^{32}P-60-nt−ssDNA produced nearly identical results with an estimated dissociation constant of 27 ± 8 nM (data not shown). The estimated dissociation constants are within the range of values (1−45 nM) previously determined for native XPC using a variety of undamaged and damaged DNA substrates (*57*, *70*, *71*).

RAD23B and CETN2 Stimulate the DNA-Binding Activity of XPC-C. Having demonstrated that XPC(492−940) can bind ssDNA and that it physically interacts with RAD23B and CETN2, we next assessed whether these proteins stimulate the ssDNA-binding activity of XPC(492−940), as observed for full-length XPC. Studies by Craescu and co-workers (*72*) demonstrated that the N-terminal extension of CETN2 mediated self-polymerization. To facilitate working with CETN2, a construct lacking the N-terminal 21 amino acids [CETN2(22−172)] was generated. A ^{15}N−^{1}H HSQC spectrum of ^{15}N-enriched-CETN2(22−172) is quite similar to the wild-type protein (Supporting Information, Figure 7), indicating the N-terminal deletion does not affect the CETN2 structure.

A mixture of XPC(492−940), RAD23B, and CETN2(22−172) in an ∼1:1.2:1.2 molar ratio was applied to a ssDNA−cellulose column, taking care that the total amount and concentration of applied XPC(492−940) was identical to that used in all the experiments. As noted above for free XPC-(492−940), the protein bound to the ssDNA−cellulose (initially 0.5 M NaCl) and was retained until elution with 2.0 M NaCl (Figure 7A). As anticipated, there was a significant increase in the amount of XPC(492−940) retained on the column in the presence of RAD23B and/or CETN2-(22−172) (compare panels A and B of Figure 7). Interestingly, CETN2(22−172) eluted in the same fractions as XPC-C and with an apparent 1:1 stoichiometry (Figure 7B).

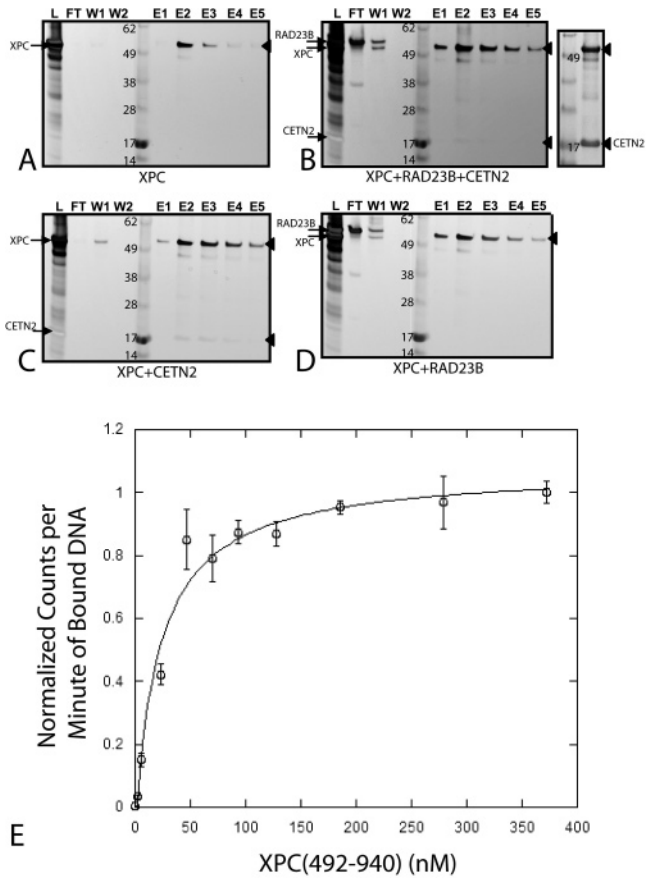

FIGURE 7: RAD23B and CETN2 stimulate ssDNA binding by XPC(492−940). (A−D) Binding of XPC(492−940) to ssDNA−cellulose columns alone (A) or the presence of RAD23B (D), CETN2(22−172) (C), or RAD23B and CETN2(22−172) (B). Sample solutions for application to the ssDNA−cellulose columns were prepared to a final volume of 5.24 mL by mixing the appropriate amounts of stock protein solutions [XPC(492−940), 3.7 μM; RAD23B, 133 μM; CETN2(22−172), 366 μM] and 0.5 M NaCl dialysate (see Experimental Procedures) to achieve the following final concentrations: XPC(492−940), 3.5 μM; RAD23B, 4.6 μM; and CETN2(22−172), 4.2 μM. The ratio of proteins in panel B was 1:1.3:1.2, respectively. SDS−PAGE (4−12%) was visualized by silver staining with a development time of 8 min. Arrowheads mark XPC and CETN2. The lanes in order are as follows: L, starting sample; FT, column flowthrough; W, wash; MW standards; E, elution. The added gel at the right in panel B shows the pooled elution fractions E1 through E5, which were concentrated 20-fold and analyzed by SDS−PAGE with Simply Blue SafeStain staining to aid in the visualization of the CETN2-(22−172) component. (E) DNA filter binding assay for XPC(492−940). XPC(492−940) was titrated against undamaged ^{32}P-60-nt−ssDNA to produce a saturation binding curve. The data points represent scintillation counts per minute from the retained XPC−^{32}P-DNA complex plotted against XPC(492−940) concentration (determined by amino acid analysis). Each data point in the assay was collected in triplicate, and the error bars represent one standard deviation from the mean.

Since CETN2(22−172) did not bind to ssDNA−cellulose in the absence of XPC(492−940), these data imply that the two proteins form a strong complex. In the case of RAD23B there was no protein retained, indicating that it did not form a stable complex with XPC(492−940).

Intrigued by stable complex formation with CETN2(22−172) but not RAD23B, and the apparent stimulation of XPC(492−940) DNA-binding activity by both (Figure 7A,B), we sought to clarify the individual contributions of

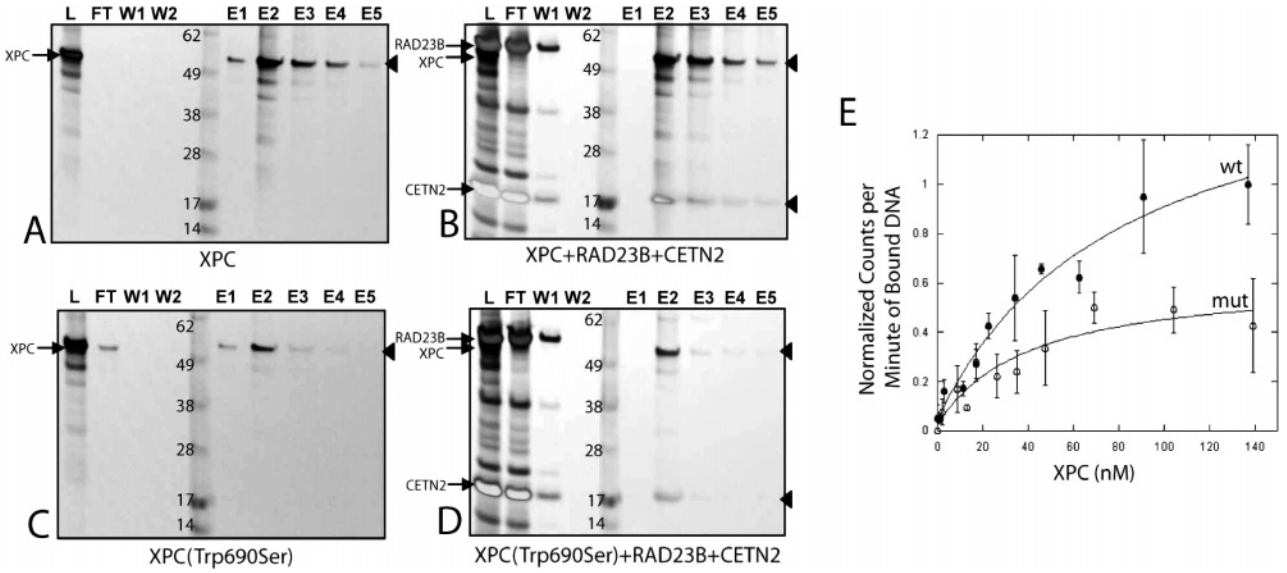

FIGURE 8: Mutation of Trp690 to Ser diminishes the DNA-binding ability of XPC. (A−D) Binding to ssDNA−cellulose columns in the absence (A, C) or the presence (B, D) of RAD23B and CETN2. Wild-type protein was used in (A) and (B) and the Trp690Ser mutant in (C) and (D). Samples for application to the columns were prepared to a final volume of 4.5 mL and the following final concentrations: XPC(492−940) (wild type and mutant), 10 μM; RAD23B, 12 μM; CETN2(22−172), 12 μM. SDS−PAGE (4−12%) was performed, and proteins were visualized by silver staining with a development time of 6 min. The order of the lanes is as described in Figure 7. Arrowheads mark XPC(492−940) and CETN2. (E) DNA filter binding assay for wild-type (filled) and Trp690Ser (open) XPC(492−940). The data points represent scintillation counts per minute from the retained damaged XPC-C−^{32}P-DNA complex plotted against XPC(492−940) concentration (determined by amino acid analysis). Each data point was collected in triplicate, and the error bars represent one standard deviation from the mean.

RAD23B and CETN2(22−172) to the stimulation of DNA-binding activity. Therefore, the DNA-binding assays were repeated in the presence of either of the XPC partners alone. These experiments showed that ssDNA binding by XPC-(492−940) was stimulated by the presence of either RAD23B (cf. Figure 7A,D) or CETN2(22−172) (cf. Figure 7A,C). Thus, each binding partner makes a contribution to the stimulation of XPC DNA binding. What is remarkable is that this is achieved in the case of CETN2 by formation of a tight complex, whereas in the case of RAD23B stimulation of DNA binding by XPC-C occurs through transient interactions. Since control experiments showed that neither RAD23B nor CETN2(22−172) binds directly to DNA, the stimulation of DNA-binding activity must be via an allosteric mechanism involving stabilization of the structure or induction of a specific conformational change in XPC (*11*).

Biochemical Basis for XP-C Defects Associated with the Trp690Ser Mutation. XPC(492−940) represents the first reported bacterially expressed construct possessing the essential DNA-binding functions of the native protein. Its availability opens the door to characterization of the biochemical properties of XPC mutants associated with XP disease. The first such analysis has been performed on the Trp690Ser mutation found in an XP-C patient described by Stefanini and colleagues (XP13PV in ref *22*). Trp690Ser is a rare missense mutation and is associated with XP-C disease; the majority of XP-C patients have nonsense mutations that truncate XPC and destroy its vital C-terminus-mediated recruitment of TFIIH (*5*, *6*, *22*). In heterozygous XP-C carriers the normal *XPC* allele usually compensates effectively for the mutated allele; however, in the patient XP13PV the *XPC* Trp690Ser paternal allele could not compensate for the defective (nonsense) maternal allele.

To investigate the biochemical consequences of the Trp690Ser mutation, an XPC(492−940) construct [XPC-C(W690S)] was generated and the protein produced using the protocols described above. Limited proteolysis experiments using proteinase K, trypsin, and chymotrypsin were performed in parallel on both wild-type and the mutant XPC-C, and identical proteolysis patterns were observed (data not shown). This implies that the mutation does not disrupt the stability or structural integrity of the XPC-C tertiary structure.

Since the Trp690Ser mutation lies within the purported DNA-binding region of XPC, XPC-C(W690S) was assessed for its ability to bind DNA using the ssDNA−cellulose and filter-binding assays. Panels A and C of Figure 8 show that, in the absence of RAD23B and CETN2, the mutant protein has a significantly weaker affinity for ssDNA than wild-type XPC(492−940) (~48% reduction in binding based on in-gel densitometry). In the presence of RAD23B and CETN2, the discrepancy in mutant binding is even more pronounced (Figure 8B,D). Thus, although there is a dramatic increase in binding of ssDNA by wild-type XPC(492−940) in the presence of these partners, ssDNA binding of XPC-C(W690S) does not appear to be stimulated by the presence of RAD23B and CETN2. CETN2 coelutes with XPC-C(W690S) as observed for wild-type XPC(492−940 (Figure 8D), confirming that the Trp690Ser mutation does not alter CETN2 binding. Moreover, RAD23B also transiently interacts with XPC-C(W690S) as observed for wild-type XPC-(492−940) in protease protection experiments (data not shown). Figure 8E summarizes NFBA experiments that show titration of increasing amounts of either wild-type XPC(492−940) or XPC-C(W690S) to undamaged ^{32}P-60-nt−ssDNA results in very different saturation binding curves: the mutanted protein retained ~50% less ^{32}P-60-nt−ssDNA than its wild-type counterpart. Thus, the Trp690Ser mutation causes a significant defect in the ssDNA-binding activity of XPC.

DISCUSSION

Specific biochemical functions of XPC have been mapped only to the C-terminal portion of the protein. However, N-terminal truncation of XPC has been shown to reduce cell-free NER activity by 65% (*6*) and diminish repair of 6-4PPs (*73*), suggesting a regulatory role for the N-terminal portion of XPC. Although interaction between XPC and XPA has been previously reported (*14*), the earlier work did not characterize further the region(s) of XPC responsible for interactions with XPA. Our discovery of a stable, structured region in the N-terminal half of XPC that physically interacts with XPA is significant because it provides the first direct evidence of a specific functional role for the N-terminal region of XPC.

While the XPC N-terminal domain displayed the strongest interaction with XPA, XPC-C was also found to physically interact with XPA, albeit considerably more weakly than XPC(154−331) (Figure 5). This observation suggests that the N-terminal domain of XPC and the C-terminal region cooperate in recruiting XPA. The coordination of XPA binding sites in XPC fits nicely with the structural modeling of the XPC core, which suggests a structural link between the N- and C-terminal portions of the protein. Further insights into the mechanism of action can be obtained by considering what is known about how XPA and XPC function.

The role for XPC−XPA interactions may be to recruit XPA to lesion sites, thereby coordinating the transition between steps in DNA repair. XPC, in association with RAD23B and CETN2, is involved in the initial damage recognition step of GG-NER (*11*). In order to transition from the damage recognition phase to the damage excision phase, a few key events must occur. These include XPC-dependent recruitment of TFIIH, opening of DNA around the lesion site, and recruitment of XPA and RPA, as well as the excision nucleases XPG and XPF−ERCC1. XPA−RPA may displace XPC (*14*, *16*) so that XPA can verify DNA damage and help to position the excision machinery (*15*−*17*).

RAD23B is required for displacement of XPC from DNA by XPA−RPA (*11*). A proposed mechanism attributes this to overlap of the RAD23B and DNA-binding sites on XPC (*11*). The binding of XPA to sites in both the N- and C-terminal portions of XPC implies that there is an XPA interaction in reasonably close proximity to the DNA- and RAD23B-binding regions on XPC. The structural model of XPC-TGC (Figure 4) suggests that the folded core brings the N- and C-terminal regions together, which would facilitate the cooperative action of the two XPA contact points and their action on XPC. An additional important observation was that XPA binds to XPC with weak to moderate affinity based on the significantly weaker strength of pull-downs relative to the strong interaction of XPC with CETN2. Affinity in this range seems highly appropriate in this context, so as to allow subsequent interactions with other essential NER factors. Our observations thus provide supporting evidence for the model of Lee and workers regarding the mechanism for XPA displacement of XPC from DNA and agree with work establishing that GG-NER proceeds via a sequential assembly mechanism rather than via a discrete repairosome (*13*, *74*).

It has been established that RAD23B forms a tight complex with XPC (*37*), providing stability (*26*), stimulation of DNA-binding activity (*10*), and protection from proteasomal degradation (*29*). There remains ambiguity, however, over the precise amino acid regions of XPC involved in RAD23B interaction. Legerski and co-workers originally used yeast two-hybrid studies to map the RAD23B-interacting region of XPC to residues 776 through 801 (*75*). Hanaoka and co-workers subsequently performed bidirectional XPC truncation studies to map the minimal RAD23B-interacting region of XPC between amino acids 496 and 734 (*6*) (Figure 1C).

Our investigations with XPC(492−940) are consistent in part with the observations of both groups. RAD23B showed transient interactions with XPC(492−940) that were capable of stimulating its DNA-binding activity (Figures 7 and 8). While we cannot exclude the possibility of inconsistencies between XPC(492−940) and native XPC because of refolding or the requirement of XPC posttranslational modifications, the transient nature of the interaction of RAD23B with XPC(492−940) suggests that there is an additional binding element in the N-terminal region of XPC. Thus, the full affinity observed for the native protein would result from cooperation of the previously mapped site in XPC-C with a second site in XPC-N. Two lines of evidence support this hypothesis. (1) In the report by Uchida et al. (2002) that mapped the RAD23B binding region of XPC (Figure 3 of that report), RAD23B weakly pulled down the XPC(496−940) truncation construct but very strongly pulled down XPC(118−940). (2) In the recent X-ray crystal structure of the complex between yeast PNGase and the XPC-binding domain of yeast Rad23 (PDB code 1X3W) the Rad23 binding surface is composed of two spatially adjacent helices that are ∼400 residues apart in the sequence, one from the N-terminal and one from the C-terminal halves of the protein (*50*). Considering the sequence similarity of XPC with PNGases and that the two subdomains were divided by a large insertion (*64*), it is possible that XPC binding to RAD23B mirrors that of the PNGase binding to yeast Rad23. Clearly, additional experiments are required to test this hypothesis.

Concluding Remarks. Significant advances have been made in understanding the DNA-damage recognition by XPC, and indeed for the entire GG-NER pathway (*16*, *24*). However, there continues to be a critical gap in knowledge regarding the biochemical mechanisms and 3D structures of individual GG-NER components and the complexes they form, especially in the case of XPC. To date, out of the total of 940 residues the only structurally characterized portion of XPC is a 17-residue fragment that binds to CETN2 (*28*, *76*). Many of the questions regarding XPC function and its domain structure, architecture, and scaffold-like ability to interact with multiple proteins as well as DNA remain to be answered. Advances toward answers to these questions have been hampered by the inability to produce significant amounts of soluble XPC. Our results using a domain-based strategy driven by sequence analysis and systematic subcloning have crossed an important threshold toward this objective.

The ability to identify and characterize the functional defect in the XPC Trp690Ser mutant associated with XP disease demonstrates the value of our approach. The mutation did not appear to result in misfolding of XPC. Rather, the mutant protein was significantly less capable of binding to

DNA, and stimulation of its DNA binding by RAD23B and CETN2 was drastically reduced. Interestingly, a mutation inserting Val at amino acid 698 has also been observed in an XP-C patient, highlighting the importance of this region in XPC function (*77*). While additional studies are necessary to elucidate the specific role of Trp690 in modulating DNA binding, knowledge of the molecular basis of disease at this level of detail offers potentially powerful approaches to develop therapeutic strategies for XP patients.

ACKNOWLEDGMENT

We thank Drs. Kevin Weiss, Craig Vanderkooi, and Jonathan Sheehan for assistance with NMR experiments, Erik Hysinger and Andrew Nager for assistance with NFBAs, Dr. David Friedman for assistance with mass spectrometry, Dr. Eric Dawson for guidance with protein modeling, and Drs. Jeffrey Salisbury and Brandt Eichman for providing plasmids. We additionally thank Drs. Craig Vanderkooi, Jonathan Sheehan, Jeffrey Salisbury, and Acu Ilie for helpful discussions and Dr. Mark Ehrhardt for critical review of the manuscript.

SUPPORTING INFORMATION AVAILABLE

Nine figures showing data as described in the text. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Sugasawa, K., Ng, J. M., Masutani, C., Iwai, S., van der Spek, P. J., Eker, A. P., Hanaoka, F., Bootsma, D., and Hoeijmakers, J. H. (1998) Xeroderma pigmentosum group C protein complex is the initiator of global genome nucleotide excision repair, *Mol. Cell 2*, 223−232.
2. Friedberg, E. C., Walker, G. C., Siede, W., Wood, R. D., Schultz, R. A., and Ellenberger, T. (2005) *DNA Repair and Mutagenesis*, 2nd ed., ASM Press, Herndon, VA.
3. Hanawalt, P. C. (2002) Subpathways of nucleotide excision repair and their regulation, *Oncogene 21*, 8949−8956.
4. Sugasawa, K., Okamoto, T., Shimizu, Y., Masutani, C., Iwai, S., and Hanaoka, F. (2001) A multistep damage recognition mechanism for global genomic nucleotide excision repair, *Genes Dev. 15*, 507−521.
5. Yokoi, M., Masutani, C., Maekawa, T., Sugasawa, K., Ohkuma, Y., and Hanaoka, F. (2000) The xeroderma pigmentosum group C protein complex XPC-HR23B plays an important role in the recruitment of transcription factor IIH to damaged DNA, *J. Biol. Chem. 275*, 9870−9875.
6. Uchida, A., Sugasawa, K., Masutani, C., Dohmae, N., Araki, M., Yokoi, M., Ohkuma, Y., and Hanaoka, F. (2002) The carboxy-terminal domain of the XPC protein plays a crucial role in nucleotide excision repair through interactions with transcription factor IIH, *DNA Repair (Amsterdam) 1*, 449−461.
7. Moggs, J. G., Yarema, K. J., Essigmann, J. M., and Wood, R. D. (1996) Analysis of incision sites produced by human cell extracts and purified proteins during nucleotide excision repair of a 1,3-intrastrand d(GpTpG)-cisplatin adduct, *J. Biol. Chem. 271*, 7177−7186.
8. Wang, G., Chuang, L., Zhang, X., Colton, S., Dombkowski, A., Reiners, J., Diakiw, A., and Xu, X. S. (2004) The initiative role of XPC protein in cisplatin DNA damaging treatment-mediated cell cycle regulation, *Nucleic Acids Res. 32*, 2231−2240.
9. Wang, G., Dombkowski, A., Chuang, L., and Xu, X. X. (2004) The involvement of XPC protein in the cisplatin DNA damaging treatment-mediated cellular response, *Cell Res. 14*, 303−314.
10. Sugasawa, K., Masutani, C., Uchida, A., Maekawa, T., van der Spek, P. J., Bootsma, D., Hoeijmakers, J. H., and Hanaoka, F. (1996) HHR23B, a human Rad23 homolog, stimulates XPC protein in nucleotide excision repair in vitro, *Mol. Cell. Biol. 16*, 4852−4861.
11. Nishi, R., Okuda, Y., Watanabe, E., Mori, T., Iwai, S., Masutani, C., Sugasawa, K., and Hanaoka, F. (2005) Centrin 2 stimulates nucleotide excision repair by interacting with xeroderma pigmentosum group C protein, *Mol. Cell. Biol. 25*, 5664−5674.
12. Bootsma, D., Kraemer, K. H., Cleaver, J. E., and Hoeijmakers, J. H. (1997) in *The Metabolic Basis of Inherited Disease* (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., Eds.) McGraw-Hill, New York.
13. Araujo, S. J., Nigg, E. A., and Wood, R. D. (2001) Strong functional interactions of TFIIH with XPC and XPG in human DNA nucleotide excision repair, without a preassembled repairosome, *Mol. Cell. Biol. 21*, 2281−2291.
14. You, J. S., Wang, M., and Lee, S. H. (2003) Biochemical analysis of the damage recognition process in nucleotide excision repair, *J. Biol. Chem. 278*, 7476−7485.
15. de Laat, W. L., Appeldoorn, E., Sugasawa, K., Weterings, E., Jaspers, N. G., and Hoeijmakers, J. H. (1998) DNA-binding polarity of human replication protein A positions nucleases in nucleotide excision repair, *Genes Dev. 12*, 2598−2609.
16. Riedl, T., Hanaoka, F., and Egly, J. M. (2003) The comings and goings of nucleotide excision repair factors on damaged DNA, *EMBO J. 22*, 5293−5303.
17. Tripsianes, K., Folkers, G., Ab, E., Das, D., Odijk, H., Jaspers, N. G., Hoeijmakers, J. H., Kaptein, R., and Boelens, R. (2005) The structure of the human ERCC1/XPF interaction domains reveals a complementary role for the two proteins in nucleotide excision repair, *Structure 13*, 1849−1858.
18. Friedberg, E. C. (2001) How nucleotide excision repair protects against cancer, *Nat. Rev. Cancer 1*, 22−33.
19. Kraemer, K. H., Lee, M. M., Andrews, A. D., and Lambert, W. C. (1994) The role of sunlight and DNA repair in melanoma and nonmelanoma skin cancer. The xeroderma pigmentosum paradigm, *Arch. Dermatol. 130*, 1018−1021.
20. Lehmann, A. R. (2003) DNA repair-deficient diseases, xeroderma pigmentosum, Cockayne syndrome and trichothiodystrophy, *Biochimie 85*, 1101−1111.
21. Magnaldo, T., and Sarasin, A. (2004) Xeroderma pigmentosum: from symptoms and genetics to gene-based skin therapy, *Cells Tissues Organs 177*, 189−198.
22. Chavanne, F., Broughton, B. C., Pietra, D., Nardo, T., Browitt, A., Lehmann, A. R., and Stefanini, M. (2000) Mutations in the XPC gene in families with xeroderma pigmentosum and consequences at the cell, protein, and transcript levels, *Cancer Res. 60*, 1974−1982.
23. Hollander, M. C., Philburn, R. T., Patterson, A. D., Velasco-Miguel, S., Friedberg, E. C., Linnoila, R. I., and Fornace, A. J., Jr. (2005) Deletion of XPC leads to lung tumors in mice and is associated with early events in human lung carcinogenesis, *Proc. Natl. Acad. Sci. U.S.A. 102*, 13200−13205.
24. Tapias, A., Auriol, J., Forget, D., Enzlin, J. H., Scharer, O. D., Coin, F., Coulombe, B., and Egly, J. M. (2004) Ordered conformational changes in damaged DNA induced by nucleotide excision repair factors, *J. Biol. Chem. 279*, 19074−19083.
25. Fan, L., Arvai, A. S., Cooper, P. K., Iwai, S., Hanaoka, F., and Tainer, J. A. (2006) Conserved XPB core structure and motifs for DNA unwinding: Implications for pathway selection of transcription or excision repair, *Mol. Cell 22*, 27−37.
26. Araki, M., Masutani, C., Takemura, M., Uchida, A., Sugasawa, K., Kondoh, J., Ohkuma, Y., and Hanaoka, F. (2001) Centrosome protein centrin 2/caltractin 1 is part of the xeroderma pigmentosum group C complex that initiates global genome nucleotide excision repair, *J. Biol. Chem. 276*, 18665−18672.
27. Popescu, A., Miron, S., Blouquit, Y., Duchambon, P., Christova, P., and Craescu, C. T. (2003) Xeroderma pigmentosum group C protein possesses a high affinity binding site to human centrin 2 and calmodulin, *J. Biol. Chem. 278*, 40252−40261.
28. Thompson, J. R., Ryan, Z. C., Salisbury, J. L., and Kumar, R. (2006) The structure of the human centrin 2-xeroderma pigmentosum group C protein complex, *J. Biol. Chem. 281*, 18746−18752.
29. Ng, J. M., Vermeulen, W., van der Horst, G. T., Bergink, S., Sugasawa, K., Vrieling, H., and Hoeijmakers, J. H. (2003) A novel regulation mechanism of DNA repair by damage-induced and RAD23-dependent stabilization of xeroderma pigmentosum group C protein, *Genes Dev. 17*, 1630−1645.
30. Schauber, C., Chen, L., Tongaonkar, P., Vega, I., Lambertson, D., Potts, W., and Madura, K. (1998) Rad23 links DNA repair to the ubiquitin/proteasome pathway, *Nature 391*, 715−718.

31. Green, C. M., and Almouzni, G. (2002) When repair meets chromatin. First in series on chromatin dynamics, *EMBO Rep. 3*, 28−33.
32. Yasuda, T., Sugasawa, K., Shimizu, Y., Iwai, S., Shiomi, T., and Hanaoka, F. (2005) Nucleosomal structure of undamaged DNA regions suppresses the non-specific DNA binding of the XPC complex, *DNA Repair (Amsterdam*) *4*, 389−395.
33. Sugasawa, K., Okuda, Y., Saijo, M., Nishi, R., Matsuda, N., Chu, G., Mori, T., Iwai, S., Tanaka, K., and Hanaoka, F. (2005) UV-induced ubiquitylation of XPC protein mediated by UV-DDB-ubiquitin ligase complex, *Cell 121*, 387−400.
34. Wang, Q. E., Wani, M. A., Chen, J., Zhu, Q., Wani, G., El-Mahdy, M. A., and Wani, A. A. (2005) Cellular ubiquitination and proteasomal functions positively modulate mammalian nucleotide excision repair, *Mol. Carcinog. 42*, 53−64.
35. Wang, Q. E., Zhu, Q., Wani, G., El-Mahdy, M. A., Li, J., and Wani, A. A. (2005) DNA repair factor XPC is modified by SUMO-1 and ubiquitin following UV irradiation, *Nucleic Acids Res. 33*, 4023−4034.
36. El-Mahdy, M. A., Zhu, Q., Wang, Q. E., Wani, G., Praetorius-Ibba, M., and Wani, A. A. (2006) Cullin 4A-mediated proteolysis of DDB2 protein at DNA damage sites regulates in vivo lesion recognition by XPC, *J. Biol. Chem. 281*, 13404−13411.
37. Masutani, C., Sugasawa, K., Yanagisawa, J., Sonoyama, T., Ui, M., Enomoto, T., Takio, K., Tanaka, K., van der Spek, P. J., Bootsma, D., et al. (1994) Purification and cloning of a nucleotide excision repair complex involving the xeroderma pigmentosum group C protein and a human homologue of yeast RAD23, *EMBO J.* 13, 1831−1843.
38. Shimizu, Y., Iwai, S., Hanaoka, F., and Sugasawa, K. (2003) Xeroderma pigmentosum group C protein interacts physically and functionally with thymine DNA glycosylase, *EMBO J. 22*, 164−173.
39. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res. 25*, 3389−3402.
40. Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., and Bairoch, A. (2003) ExPASy: The proteomics server for in-depth protein knowledge and analysis, *Nucleic Acids Res. 31*, 3784−3788.
41. Notredame, C., Higgins, D. G., and Heringa, J. (2000) T-COFFEE: A novel method for fast and accurate multiple sequence alignment, *J. Mol. Biol. 302*, 205−217.
42. Pollastri, G., Przybylski, D., Rost, B., and Baldi, P. (2002) Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles, *Proteins 47*, 228−235.
43. Blom, N., Gammeltoft, S., and Brunak, S. (1999) Sequence and structure-based prediction of eukaryotic protein phosphorylation sites, *J. Mol. Biol. 294*, 1351−1362.
44. Ward, J. J., McGuffin, L. J., Bryson, K., Buxton, B. F., and Jones, D. T. (2004) The DISOPRED server for the prediction of protein disorder, *Bioinformatics 20*, 2138−2139.
45. Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F., and Jones, D. T. (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life, *J. Mol. Biol. 337*, 635−645.
46. Dosztanyi, Z., Csizmok, V., Tompa, P., and Simon, I. (2005) The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins, *J. Mol. Biol. 347*, 827−839.
47. Masutani, C., Araki, M., Sugasawa, K., van der Spek, P. J., Yamada, A., Uchida, A., Maekawa, T., Bootsma, D., Hoeijmakers, J. H., and Hanaoka, F. (1997) Identification and characterization of XPC-binding domain of hHR23B, *Mol. Cell. Biol. 17*, 6915−6923.
48. Alberts, B., and Herrick, G. (1971) DNA-cellulose chromatography. *Methods Enzymol. 21*, 198−217.
49. McEntee, K., Weinstock, G. M., and Lehman, I. R. (1980) recA protein-catalyzed strand assimilation: stimulation by *Escherichia coli* single-stranded DNA-binding protein, *Proc. Natl. Acad. Sci. U.S.A. 77*, 857−861.
50. Lee, J. H., Choi, J. M., Lee, C., Yi, K. J., and Cho, Y. (2005) Structure of a peptide:N-glycanase-Rad23 complex: insight into the deglycosylation for denatured glycoproteins, *Proc. Natl. Acad. Sci. U.S.A. 102*, 9144−9149.
51. Sali, A., and Blundell, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints, *J. Mol. Biol. 234*, 779−815.
52. Soding, J., Biegert, A., and Lupas, A. N. (2005) The HHpred interactive server for protein homology detection and structure prediction, *Nucleic Acids Res. 33*, W244−W248.
53. Canutescu, A. A., Shelenkov, A. A., and Dunbrack, R. L., Jr. (2003) A graph-theory algorithm for rapid protein side-chain prediction, *Protein Sci. 12*, 2001−2014.
54. Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham, T. E., III, DeBolt, S., Ferguson, D., Seibel, G., and Kollman, P. (1995) AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules, *Comput. Phys. Commun. 91*, 1−41.
55. Morris, A. L., MacArthur, M. W., Hutchinson, E. G., and Thornton, J. M. (1992) Stereochemical quality of protein structure coordinates, *Proteins 12*, 345−364.
56. Gregoret, L. M., and Cohen, F. E. (1990) Novel method for the rapid evaluation of packing in protein structures, *J. Mol. Biol. 211*, 959−974.
57. Reardon, J. T., Mu, D., and Sancar, A. (1996) Overproduction, purification, and characterization of the XPC subunit of the human DNA repair excision nuclease, *J. Biol. Chem. 271*, 19451−19456.
58. Kamionka, M., and Feigon, J. (2004) Structure of the XPC binding domain of hHR23A reveals hydrophobic patches for protein interaction, *Protein Sci. 13*, 2370−2377.
59. Siede, W., and Eckardt-Schupp, F. (1986) DNA repair genes of *Saccharomyces cerevisiae*: complementing rad4 and rev2 mutations by plasmids which cannot be propagated in *Escherichia coli*, *Curr. Genet. 11*, 205−210.
60. Wei, S., and Friedberg, E. C. (1998) A fragment of the yeast DNA repair protein Rad4 confers toxicity to E. coli and is required for its interaction with Rad7 protein, *Mutat. Res. 400*, 127−133.
61. Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L., Studholme, D. J., Yeats, C., and Eddy, S. R. (2004) The Pfam protein families database, *Nucleic Acids Res. 32*, D138−D141.
62. Marti, T. M., Kunz, C., and Fleck, O. (2003) Repair of damaged and mismatched DNA by the XPC homologues Rhp41 and Rhp42 of fission yeast, *Genetics 164*, 457−467.
63. Tsumoto, K., Ejima, D., Kumagai, I., and Arakawa, T. (2003) Practical considerations in refolding proteins from inclusion bodies, *Protein Expression Purif. 28*, 1−8.
64. Anantharaman, V., Koonin, E. V., and Aravind, L. (2001) Peptide-*N*-glycanases and DNA repair proteins, Xp-C/Rad4, are, respectively, active and inactivated enzymes sharing a common transglutaminase fold, *Hum. Mol. Genet. 10*, 1627−1630.
65. Ginalski, K., Elofsson, A., Fischer, D., and Rychlewski, L. (2003) 3D-Jury: a simple approach to improve protein structure predictions, *Bioinformatics 19*, 1015−1018.
66. Thoma, B. S., Wakasugi, M., Christensen, J., Reddy, M. C., and Vasquez, K. M. (2005) Human XPC-hHR23B interacts with XPA-RPA in the recognition of triplex-directed psoralen DNA interstrand crosslinks, *Nucleic Acids Res. 33*, 2993−3001.
67. Shivji, M. K., Eker, A. P., and Wood, R. D. (1994) DNA repair defect in xeroderma pigmentosum group C and complementing factor from HeLa cells, *J. Biol. Chem. 269*, 22749−22757.
68. Wakasugi, M., and Sancar, A. (1998) Assembly, subunit composition, and footprint of human DNA repair excision nuclease, *Proc. Natl. Acad. Sci. U.S.A. 95*, 6669−6674.
69. Sugasawa, K., Shimizu, Y., Iwai, S., and Hanaoka, F. (2002) A molecular mechanism for DNA damage recognition by the xeroderma pigmentosum group C protein complex, *DNA Repair (Amsterdam*) *1*, 95−107.
70. Hey, T., Lipps, G., Sugasawa, K., Iwai, S., Hanaoka, F., and Krauss, G. (2002) The XPC-HR23B complex displays high affinity and specificity for damaged DNA in a true-equilibrium fluorescence assay, *Biochemistry 41*, 6583−6587.
71. Trego, K. S., and Turchi, J. J. (2006) Pre-steady-state binding of damaged DNA by XPC-hHR23B reveals a kinetic mechanism for damage discrimination, *Biochemistry 45*, 1961−1969.
72. Tourbez, M., Firanescu, C., Yang, A., Unipan, L., Duchambon, P., Blouquit, Y., and Craescu, C. T. (2004) Calcium-dependent self-assembly of human centrin 2, *J. Biol. Chem. 279*, 47672−47680.
73. Emmert, S., Kobayashi, N., Khan, S. G., and Kraemer, K. H. (2000) The xeroderma pigmentosum group C gene leads to selective repair of cyclobutane pyrimidine dimers rather than 6-4 photoproducts, *Proc. Natl. Acad. Sci. U.S.A. 97*, 2151−2156.

74. Volker, M., Mone, M. J., Karmakar, P., van Hoffen, A., Schul, W., Vermeulen, W., Hoeijmakers, J. H., van Driel, R., van Zeeland, A. A., and Mullenders, L. H. (2001) Sequential assembly of the nucleotide excision repair factors in vivo, *Mol. Cell 8*, 213−224.
75. Li, L., Lu, X., Peterson, C., and Legerski, R. (1997) XPC interacts with both HHR23B and HHR23A in vivo, *Mutat. Res. 383*, 197−203.
76. Yang, A., Miron, S., Mouawad, L., Duchambon, P., Blouquit, Y., and Craescu, C. T. (2006) Flexibility and plasticity of human centrin 2 binding to the xeroderma pigmentosum group C protein (XPC) from nuclear excision repair, *Biochemistry 45*, 3653−3663.
77. Li, L., Bales, E. S., Peterson, C. A., and Legerski, R. J. (1993) Characterization of molecular defects in xeroderma pigmentosum group C, *Nat. Genet. 5*, 413−417.
78. Nicholls, A., Sharp, K. A., and Honig, B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons, *Proteins 11*, 281−296.
79. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera—a visualization system for exploratory research and analysis, *J. Comput. Chem. 25*, 1605−1612.

BI061370O