PNAS
www.pnas.org

Supplement for

Human XPG nuclease structure, assembly and activities with insights for neurodegeneration and cancer from pathogenic mutations

Susan E. Tsutakawa, Altaf H. Sarker, Clifford Ng, Andrew S. Arvai, David S. Shin, Brian Shih, Shuai Jiang, Aye C. Thwin, Miaw-Sheue Tsai, Alexandra Willcox, Mai Zong Her, Kelly S. Trego, Alan G. Raetz, Daniel Rosenberg, Albino Bacolla, Michal Hammel, Jack D. Griffith*, Priscilla K. Cooper*, and John A. Tainer*.

* *Co-corresponding Authors* John Tainer, Priscilla Cooper, Jack Griffith, and Susan Tsutakawa
Email: jatainer@gmail.com, pkcooper@lbl.gov, jdg@med.unc.edu, setsutakawa@lbl.gov

Supplementary materials for this manuscript include the following:

Online Methods
Supplemental References
Figures S1 to S7
Tables S1 to S4

www.pnas.org/cgi/doi/10.1073/pnas.1921311117

ONLINE METHODS

Antibodies Antibodies used in this study include α-GAPDH (Millipore), α -tubulin (Calbiochem), α-Flag (Sigma), XPG antibody R2 # 97727 (1) and a polyclonal rabbit α-XPGcat that we had made for this study.

Protein Purification. The human catalytic domain was cloned with residues 79–785 replaced with residues 89–128 from *P. furiosus* FEN-1 (2) and with a C-terminal truncation after residue 987. XPGcat was transformed into Rosetta2 DE3 cells, expression was induced with 1 mM IPTG at 0.8 OD. Cells were harvested after overnight at 16°C. The catalytic domain was purified by Q, SP, Heparin and size exclusion chromatography (SEC). The protein was concentrated and stored in 25 mM Tris, pH 7.5, 200 mM KCl, 1 mM EDTA, 10% glycerol, 1 mM DTT, 1 EDTA-free Roche Tablet/liter.

XPGFLAG

XPGΔR and untagged full-length XPG were purified as previously described (1). Full-length XPG with a C-terminal FLAG tag was expressed in High Five insect cells. Cells were harvested 48 hours after infection and purified by anti-FLAG M2 resin (Sigma), SP sepharose and Superose 6. The protein was stored in 20 mM Tris, pH 7.5, 100 mM KCl, 5 mM $CaCl_2$, 1 mM EDTA, 1 mM 2-mercaptoethanol (BME), 10% glycerol and 0.1% OG or in the case of SAXS, immediately analyzed.

Incision assays. XPG is product inhibited, and protein was assayed until single turnover questions with enzyme excess. The DNA substrates were formed as outlined in Table S2. The ends of the 15 nt bubble DNA were blocked with a 9 nt hairpin. 5'-^{32}P -end labeled 15-nt (100 fmol: 3.3 nM) bubble substrates were incubated at 37°C with (33.3 nM) of untagged XPG, WTXPG-FLAG or mutant version of XPG-FLAG (as indicated in figures) in a buffer (25 mM Tris, pH 6.8, 10 % glycerol, 2.5 mM BME, 0.5 mg/ml BSA, 30 mM KCl and 4 mM $MgCl_2$) in 15 µl reaction volume for 15, 30, 60 or 90 min. The Y-substrate was formed by annealing 5'-^{32}P -end labeled 5T-20 and 20-10C. 10 nM substrates were incubated with 25 nM XPGcat or various mutants of XPGcat at 16 °C for 15, 30, 45 and 60 sec. in a buffer containing 25 mM Tris, pH 6.8, 10% glycerol, 2.5 mM BME, 0.5 mg/ml BSA, 0.3 mM EDTA, 30 mM KCl and 2 mM $MgCl_2$. After the incubation the reaction mixtures were stopped with formamide loading buffer (90% formamide, 0.5x TBE, 0.1% bromophenol blue and xylene cyanol FF) and were separated by 12% denaturing PAGE. The experiments were conducted three times and the bands were visualized by STORM imaging system and quantitated by Molecular Dynamics (ImageQuant) or visualized and quantified by Versadoc system (BioRad). Assays were done in triplicate, on different days.

Electrophoretic Mobility Shift Analysis (EMSA) For DNA binding of XPGcat or mutant constructs. 5 nM of the 5'-^{32}P-end labeled bubble or Y-substrates were incubated with proteins as indicated in the figure in a binding buffer containing 10 mM HEPES pH 7.5, 110 mM KCl, 1 mM EDTA, 1 mM DTT, 10 mM $CaCl_2$, 100 ng of poly dI-dC and 4% glycerol at room temperature for 15 min. For DNA binding of WT XPG or various mutants, all the above binding conditions are the same except 10 mM $CaCl_2$ was replaced with 0.1% NP-40. The reaction mixtures were separated by 4.5% native gel in cold room and visualized by STORM imaging system and quantitated by Molecular Dynamics (ImageQuant) or visualized and quantified by Versadoc system (BioRad). Assays were done in triplicate, on different days.

Crystallization and Structure Determination. Data from three crystals of human XPGcat were merged for structure refinement. XPGcat (27 mg/ml) was incubated 1:1 in one of three mother liquors: 1) 40% $AmSO_4$, 200 mM Imidizole/Malate Buffer pH 4.2,100 mM $MgCl_2$; 2) 24% $AmSO_4$, 200 mM Imidizole/Malate Buffer pH 4.2, 250 mM $MgCl_2$, 0.5 mM $SmSO_4$, 10 mM DTT; 3) 32% $AmSO_4$, 10 mM DTT, 200 mM Imidizole/Malate Buffer pH 4.2, and 50 mM $MgCl_2$. Crystals formed in hanging drops after 3-6 days at 15°C. Data was collected at 100K at beamlines 12.3.1 (ALS) for

crystal 1 or at 11-1 (SSRL) for crystals 2 and 3, with the wavelength at 1.1 Å. All three crystals diffracted to 2.1 Å, and data was indexed together. The crystal diffraction was highly anisotropic, which led to unusually high rejection of reflections(3). The resolution chosen was based on quality of the electron density in unmodeled sections. The phases were solved by molecular replacement using PHASER. The molecular replacement model was created by removing loops and overlaying the FEN chains in the following pdb files: 3Q8K, 1A76, 1B43, 1MC8, 1RXV, 1UL1, and 2IZO. The model was refined using PHENIX. Electron density for loop 110-115 was more visible in the merged data than from any single crystal. Structures were drawn using PYMOL (Schrödinger).

Biochemical analysis of multimerization. XPG-FLAG and XPG-GFP (full-length XPG with either a C-terminal FLAG or GFP tag) were expressed together or separately in High Five insect cells. Cells were harvested 48 hours after infection. When we expressed the differently tagged XPG separately, we mixed the cells post-expression. Cells were lysed by Dounce homogenizer in TBS pH 7.4, 0.5% NP-40, 1 mM PMSF, Roche EDTA-free protease inhibitors, 1 mM BME, 1 mM $MgCl_2$, 0.5 μl turbonuclease/10 ml cells. XPG-FLAG was purified from the lysate by anti-FLAG M2 resin (Sigma). Beads were washed, and XPG-FLAG was eluted by addition of (200 ng/μl 3x-FLAG peptide. The eluant from the anti-FLAG resin was sequentially incubated with beads attached to GFP binder nanobodies (4). GFP binder beads were washed. At the end, both anti-FLAG and GFP binder beads were incubated with 1x SDS sample buffer to remove bound protein. FLAG-peptide-eluted protein, protein remaining on the anti-FLAG resin, protein that did not bind to the subsequent GFP binder beads, and protein that bound the GFP binder beads were then analyzed by PAGE and Western analysis.

SAXS analysis of XPGcat, XPGΔR, and full-length XPG. SAXS data on freshly purified full-length XPG-FLAG (0.2 mg/ml) in 100 mM Tris, pH 7.5, 300 mM KCl, 0.1% OG, 1% glycerol at 10 °C was collected at beamline 12.3.1 at the Advanced Light Source synchrotron on a MarCCD 165 detector. The sample to detector distance was 1.6 m. Data was collected at 12 keV. The exposure was 5 secs with mixing in the sample cell. There was no sign of radiation damage compared to the first 0.5 sec exposure. The Rg was 73 Å, I_0 was 66 detector units. The SAXS Molecular weight estimate was 253 kD, within 20% error of a dimer based on a theoretical mass of 133 kD for a monomer. SAXS analysis was done using the program SCÅTTER (https://bl1231.als.lbl.gov/scatter/). SAXS data was also collected on XPGΔC-FLAG, XPGΔR, and XPGcat at the SIBYLS beamline. All proteins were purified by SEC just prior to SAXS data collection. The SEC buffer was specific to each protein construct. The SEC buffer for XPGΔC was the same as full-length XPG. The SEC buffer for XPGcat was 50 mM HEPES pH 7.8, 50 mM KCl, 5% glycerol, 1 mM DTT, 5 mM $CaCl_2$, 1 mM EDTA. The SEC buffer for XPGΔR was 25 mM Tris pH 7.5, 1 mM EDTA, 5 mM $CaCl_2$, 5% glycerol, 500 mM KCl, 2 mM DTT. XPGΔR was also purified bound to a 10 nt DNA with 14 bp arms on either side (Table S2); the SEC buffer was the same except including only 100 mM KCl. SAXS data collection and data analysis details are provided in Table S3.

Preparation of DNA substrates for EM. The substrate resembling a replication fork was created as previously done (5). Briefly, a single replication fork was generated at a defined site on a circular 3 kb plasmid DNA using a combination of a site-specific nicking site on the plasmid followed by a 400 bp G-less cassette which allows displacement of a 400 nt ss tail in a reaction catalyzed by DNA polymerase I (exo minus) which begins at the nick and terminates at the end of the G-less cassette if dGTP is not present in the reaction. The ss tail is converted to duplex by annealing a primer followed by synthesis to the end of the displaced strand (see details in Subramanian and Griffith 2005). For the DNA used here a 15 nt gap was present at the fork on the displaced strand. The bubble DNA substrate consisted of a 300 bp DNA constructed to have a 10 base unpaired bubble at the exact center (Fig. 4A). The DNA was generated by annealing and ligating four oligo sequences (Table S2). Briefly, 100 nt EM bubble top-strand containing 10 T at the middle were annealed with

100 nt EM bubble bottom-strand containing 10 C sequences in an annealing buffer (10 mM Tris 7.5 and 50 mM NaCl) by heating at 95°C for 3 min in a water bath followed by slowly cooling to room temperature. They are constructed in a way with 12 nt overhang after annealing for subsequent ligation with the arm. Each of 100 nt EM Arm top and bottom-strands were annealed as above and were ligated with the EM bubble over night at 16 0C with T4 ligase (NEB). The arms were designed in a way so they can ligate in both sides of the bubble. The 300 bp ligated dsDNA DNA containing 10 nt bubble at the center of the sequence were purified by Qiagen kit after electrophoresis in 2% agarose gel.

Preparation of XPG/DNA protein complexes for EM visualization. XPG purified through SEC was incubated with DNA in a buffer consisting of 10 mM HEPES pH 7.5, 110 mM KCl, 1 mM EDTA, and 4% glycerol. Incubations were carried out in 20 to 50 µl volumes using 50 ng of DNA and an amount of protein determined by experimentation to result in at least 50% of the DNA substrate being bound, while at the same time minimizing the fraction of DNA in aggregates. For most experiments these reactions contained the DNA at 2.5 µg/ml and XPG at 10 to 15 µg/ml equivalent to a ratio of 3 to 4 XPG dimers per 300 bp DNA. The DNA was pre-heated to 55°C for 5 min to break up any DNA aggregates and then incubated with XPG for 15 min at room temperature. An equal volume of 1.2% glutaraldehyde was added for 5 min at room temperature followed by chromatography over a 2 ml column of 5% agarose beads (Agarose Bead Technology Inc Burgos, Spain) equilibrated in 10 mM Tris pH 7.5, 0.1 mM EDTA. Fractions containing the XPG/DNA complexes were then prepared for EM.

Preparation of samples for EM visualization. Samples were prepared for EM analysis as previously described (6). In brief, the samples were adsorbed to thin carbon foils treated with a glow discharge in the presence of a buffer containing 2 mM spermidine and then washed through a graded water-ethanol series to 100% ethanol, air dried and rotary shadowcast with tungsten at a low angle. The samples were examined in a Tecnai T-12 instrument at 40 kV and images captured using a Gatan Orius CCD camera. Images were cropped and contrast adjusted and arranged into panels using Adobe Photoshop software. Bending angles were measured on the micrographs using Image J software and lengths measured with Gatan Digital Image software.

2-Aminopurine experiments. 6 uM 20-nt bubble DNA with 2AP at the 3'-junction was incubated alone, with 6 uM untagged XPG protein, or protein and 10 mM CaCl2. The oligonucleotide containing 2AP was used for the ssDNA control. Buffer contained 10 mM HEPES (7.5), 50 mM KCl, 1 mM EDTA at room temperature for 20 min followed by analysis by TECAN, with the following collection parameters. Excitation: 310 nm. Emission: 330-390. Integration time: 0.2 sec. 1 nm wavelength step. Slits: excitation = 5 nm / emission = 5 nm. No polarizers. Experiment was done in triplicate

Cell Culture and Recombinant XPG Expression. WI-38 normal human fetal lung fibroblasts, VA13 (SV40-transformed WI-38) fibroblasts, XP-G fibroblasts (XP125LO and XP65BE), and XP-G/CS fibroblasts (XP2BI, XPCS1LV, XPCS2LV, and XPCS1RO) were from Coriell (Camden, New Jersey). XP-G/CS primary fibroblasts XP72MA, XP165MA, and XP40GO were a gift of Steffen Emmert (University of Rostock) and were immortalized in the Cooper lab by infection with an hTERT-expressing retrovirus. Immortalized WT VA13 or various XP-G or XP-G/CS lines were cultured in DMEM containing 10% FBS, 2 mM L-glutamine, 100 U/ml penicillin and 100 µg/ml streptomycin (Life Technologies). Cell lines were lysed in RIPA buffer (10 mM Tris, pH 8.0, 140 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS) on ice and shearing the DNA with a needle connected to a syringe followed by centrifugation at 12000 rpm for 15 min at 4°C. The resulting supernatants were collected and frozen at -80°C or used immediately. For western blotting, extracts were separated using Novex 4–12% Tris–glycine gels (Invitrogen), and transfer of the protein was performed onto nitrocellulose membranes at 4°C at 35 V for overnight. Blocking was performed in 5% non-fat milk for 1 h at room temperature followed by incubation with rabbit anti-

XPG 97727 antibody (R2 at 1:1000) (1). The membrane was reprobed with anti-GAPDH antibody (Millipore, MAB374, 1:5000) or anti-Actin (Calbiochem, Ab-1, 1:5000) to obtain loading controls. Secondary antibodies used were HRP-conjugated anti-mouse and anti-rabbit and detected using the ECL Prime system (Amersham). Westerns were done twice.

For recombinant expression of XPG in XP-G/CS cells lacking endogenous XPG protein, full-length WT XPG cDNA with a C-terminal eGFP tag (https://www.addgene.org/29772/) was cloned into the pcDNA3 expression vector. Site-directed mutagenesis was performed to generate Alanine and Glutamate dimer mutants. As a "vector only" control, eGFP with an added NLS was cloned into pcDNA3. The vectors were transfected into SV40-immortalized XP-G/CS cells XPCS1RO using nucleofector technology (Lonza). Western analysis for XPG was performed 48 hrs. after transfection.

Supplemental References

1. Sarker AH*, et al.* (2005) Recognition of RNA polymerase II and transcription bubbles by XPG, CSB, and TFIIH: insights for transcription-coupled repair and Cockayne Syndrome. *Mol Cell* 20(2):187-198.
2. Hosfield DJ, Mol CD, Shen B, & Tainer JA (1998) Structure of the DNA repair and replication endonuclease and exonuclease FEN-1: coupling DNA and PCNA binding to FEN-1 activity. *Cell* 95(1):135-146.
3. Read RJ (1999) Detecting outliers in non-redundant diffraction data. *Acta Crystallogr D Biol Crystallogr* 55(Pt 10):1759-1764.
4. Meek K, Lees-Miller SP, & Modesti M (2012) N-terminal constraint activates the catalytic subunit of the DNA-dependent protein kinase in the absence of DNA or Ku. *Nucleic Acids Res* 40(7):2964-2973.
5. Subramanian D & Griffith JD (2005) p53 Monitors replication fork regression by binding to "chickenfoot" intermediates. *J Biol Chem* 280(52):42568-42572.
6. Griffith JD & Christiansen G (1978) Electron microscope visualization of chromatin and other DNA-protein complexes. *Annu Rev Biophys Bioeng* 7:19-35.

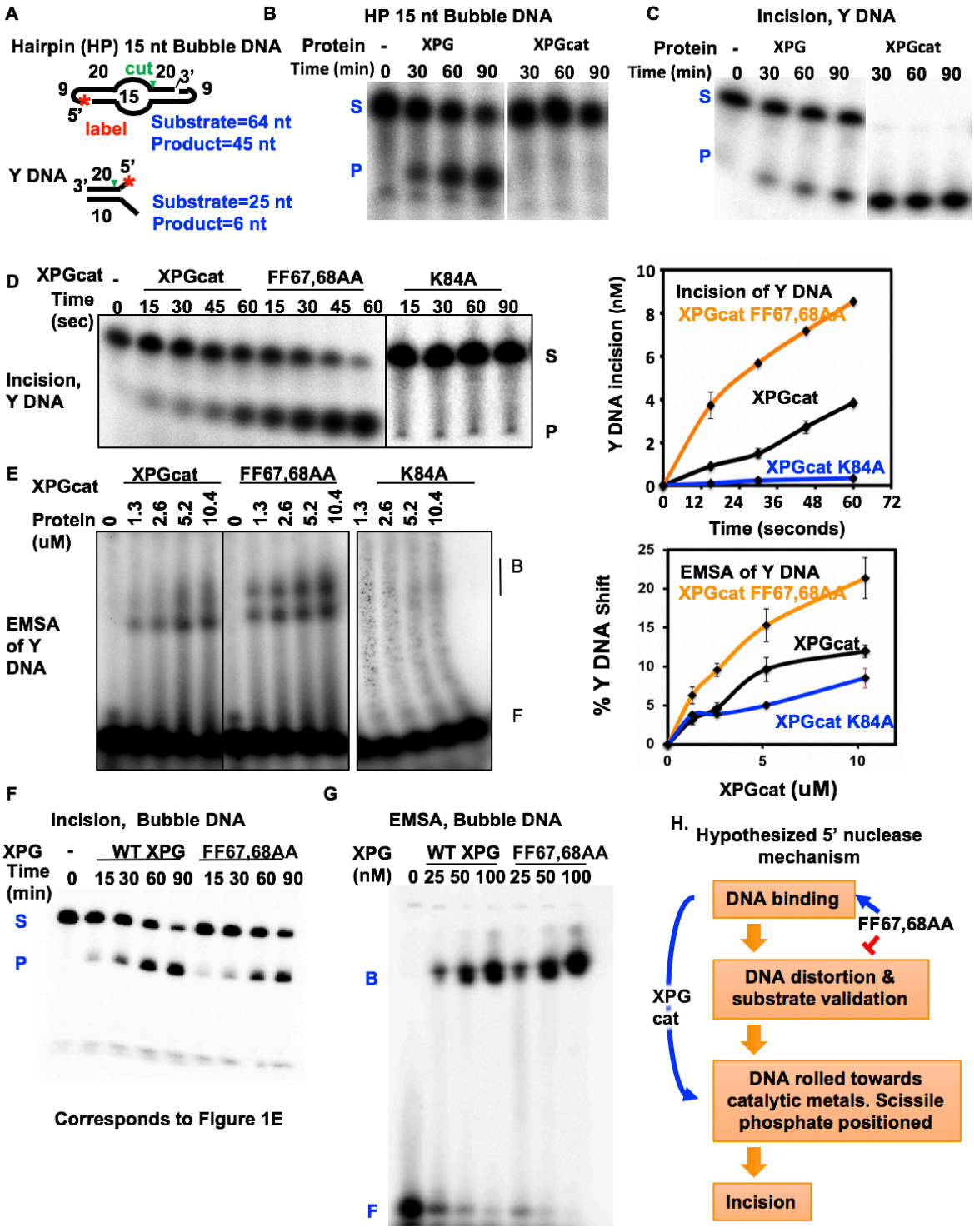


Fig. S1. Incision and DNA binding capabilities of full-length XPG and XPGcat. Denaturing or native PAGE gels are shown to indicate respectively, incision or DNA binding. For incision, substrate (S) and product (P) bands are indicated, while for EMSA, DNA-bound (B) or DNA-free (F) are marked. **A**) Schematic of ^{32}P-labelled (*) DNA substrates used in this figure. **B**) Denaturing PAGE gel showing incision activity of WT XPG-Flag on a 15 nt bubble DNA (HP15T/C) and no incision by XPGcat. **C**) Denaturing PAGE gel showing incision activity of both WT XPG and XPGcat on Y DNA.

This assay confirms that XPGcat used in crystallization studies had Y DNA incision activity. **D**) Denaturing PAGE gel showing incision activity of WT and mutant XPGcat on Y DNA. Quantitation (right) shows that the FF87,88AA mutant had increased activity against Y DNA. Mutation of Lys84, a FEN superfamily-conserved active site residue to alanine eliminated XPGcat activity. Incision activity was done in triplicate, with error bars showing standard deviation. **E**) Native PAGE gel showing WT and mutant XPGcat binding to Y DNA. Quantitation (right side) shows that the FF87,88AA mutant had significantly increased affinity, potentially explaining the increased activity in Fig. S1D. Although K84A had slightly reduced binding affinity for Y DNA, it was not enough to explain the loss of catalytic activity, consistent with Lys84 playing a catalytic role and not substrate-binding function. EMSA was done in triplicate, with error bars showing standard deviation. **F**) Representative denaturing PAGE showing single turnover incision activity of WT XPG and FF67.68AA mutant on 15 nt DNA bubble, corresponding to quantitation in Fig. 1E. **G**) Native PAGE showing DNA binding by WT XPG and FF67.68AA mutant on 15 nt DNA bubble, corresponding to quantitation in Fig. 1F. **H**) Schematic of model for 5' nuclease mechanism. Incision activity in 5' nucleases are controlled by substrate validation through substrate DNA distortion after DNA binding. Only after substrate is validated, disorder-to-order transitions in the protein repositions the scissile phosphate within catalytic distance of the metals through rolling of the duplex DNA into the active site, which licenses the DNA for incision. We postulate that Phe67 and Phe68 are involved in DNA distortion and substrate validation in full-length XPG. Thus, removal of these phenylalanines in FF67,68AA mutant increases DNA binding through reduced DNA distortion but then interferes with substrate validation. However, in XPGcat, the substrate validation step is bypassed because in the thermophilic archaeal FEN insert, the gateway is ordered and the scissile phosphate can move into catalytic position. Increased DNA binding thus leads to increased incision.

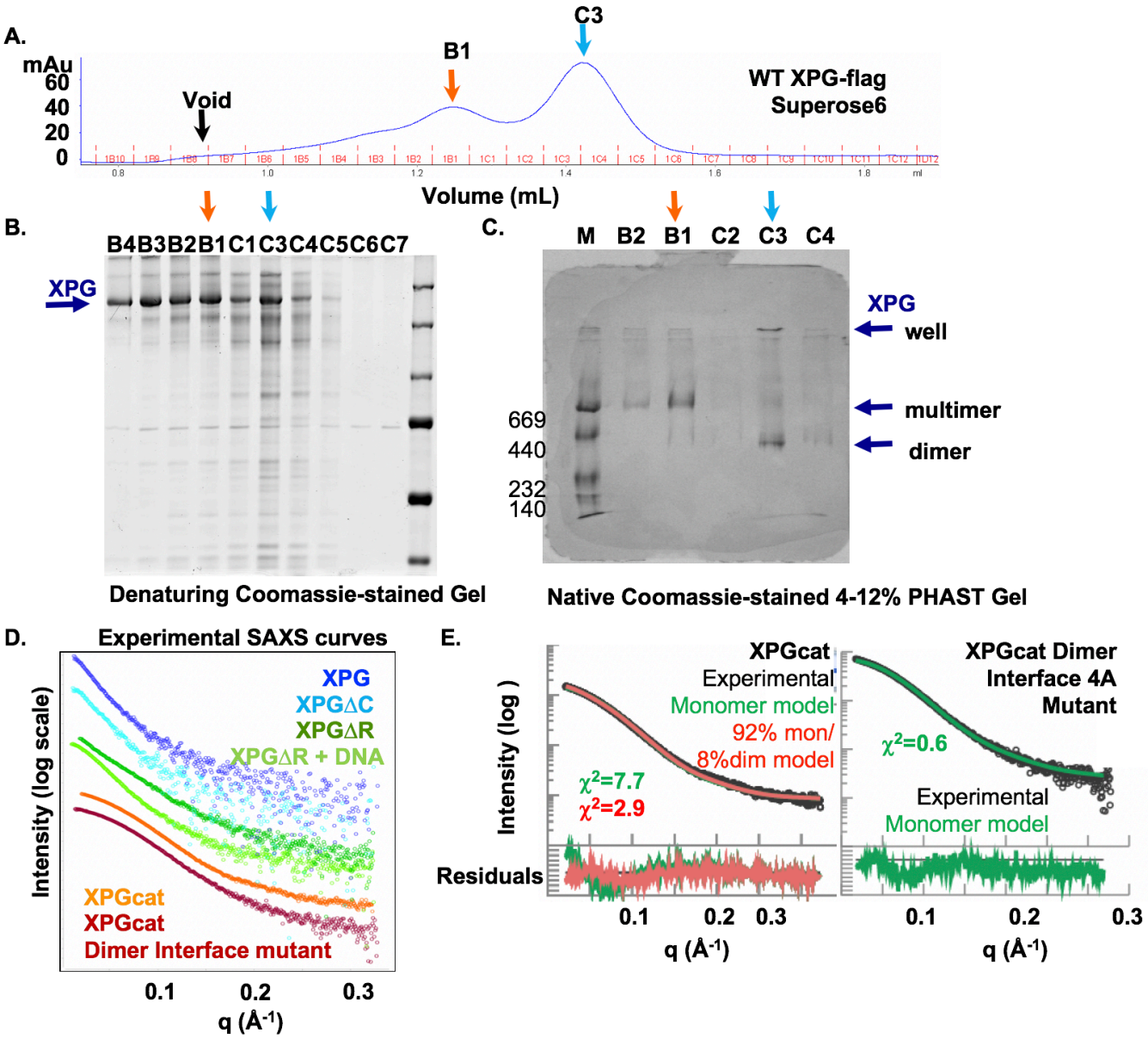
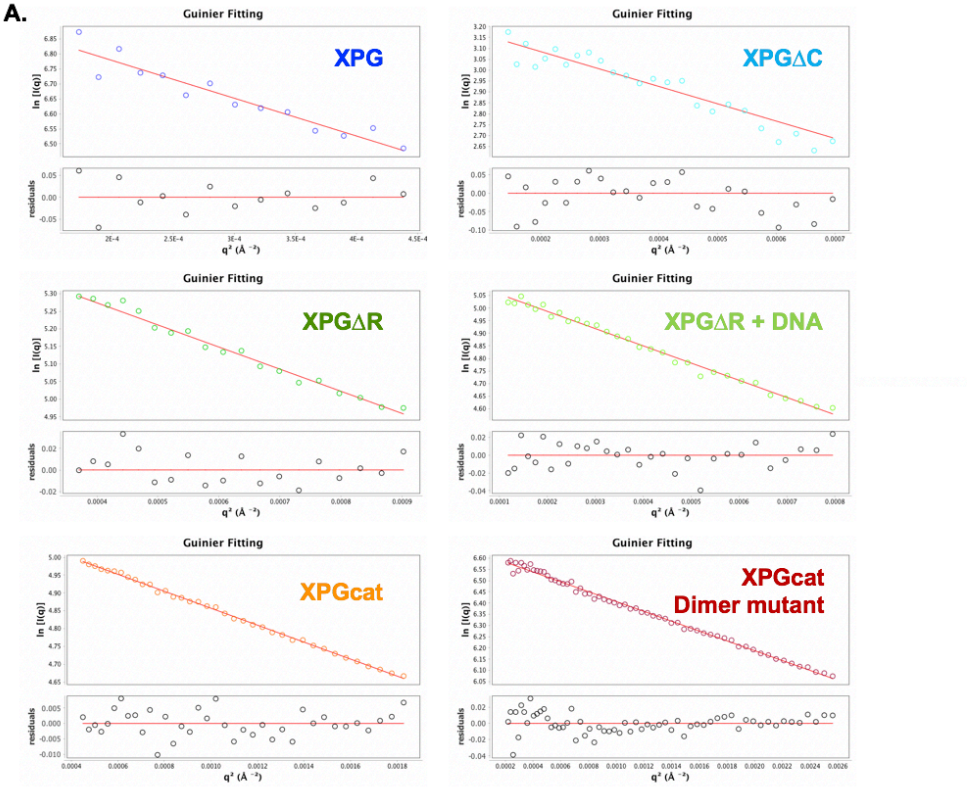


Fig. S2. SAXS analysis of full-length, truncated, and mutant XPG. A) Elution profile of WT XPG-FLAG from 2.4 mL Superose 6 column. Y axis is in-line UV absorbance at 280 nm, and x-axis is volume eluted. Fraction names are indicated along X-axis. Peaks assigned as dimer or large multimer are marked in blue and orange, respectively. No monomer peak was detected. Corresponding denaturing PAGE (**B**) and native PHAST 4-15% PAGE (**C**) gel of eluting fractions are shown. The bands below the main XPG peak were likely nicks in the XPG multimer, as the native gel shows a single band. The mobility in the native gel are consistent with a different molecular mass of the dimer and large multimer. **D**) Reciprocal space SAXS curves of full-length, truncated, and mutant XPG. XPGΔR was tested with a 10 nt bubble DNA. **E**) Fit of the WT and mutant XPGcat experimental data to the crystal structure (monomer model) or the dimer model based on the crystallographic lattice. The best fit to XPGcat included a small fraction of dimer in a minimal ensemble analysis, while the best fit to the dimer mutant 4A XPGcat did not require inclusion of any dimer. The residuals plotted below the scattering curves shows the difference between the experimental scattering curve and the predicted model scattering curve. When the residuals are not biased, the residuals should fluctuate without bias around a flat line.

A.



B.

SAXS	Calc. MW prot/DNA (kD)	SAXS MW (kD)	Mono-mer	Dimer
XPG-flag	134	265	0%	100%
XPGΔC	114	236	0%	100%
XPGΔR	62	57	91%	9%
XPGΔR+DNA	62/23	116	50%	50%
XPGcat	40	37	92%	8%
XPGcat dimer 4A mutant	40	37	100%	0%

C.

MALS	Calc. MW prot/DNA (kD)	MALS MW (kD)	Mono-mer	Dimer
XPGΔR	62	97	30%	70%
XPGΔR+DNA	62/23	153.4	0%	100%

Fig. S3. Analysis of the SAXS data of full-length, truncated, and mutant XPG. A) Guinier analysis of SAXS data shows that the Guinier regions are linear, indicating that the protein samples had no significant amounts of large aggregation. **B**) Molecular mass analysis based on SAXS data indicates the relative oligomerization state of the proteins in solution. The molecular masses of full-length XPG and XPGΔC were consistent with a dimer, XPGΔR with a monomer-dimer transition where the dimer was stabilized by a 14 nt bubble DNA, XPGcat with a monomer with a small amount of dimer, and XPGcat dimer mutant 4A with a monomer. The XPGcat dimer mutant 4A was stable and well-folded.

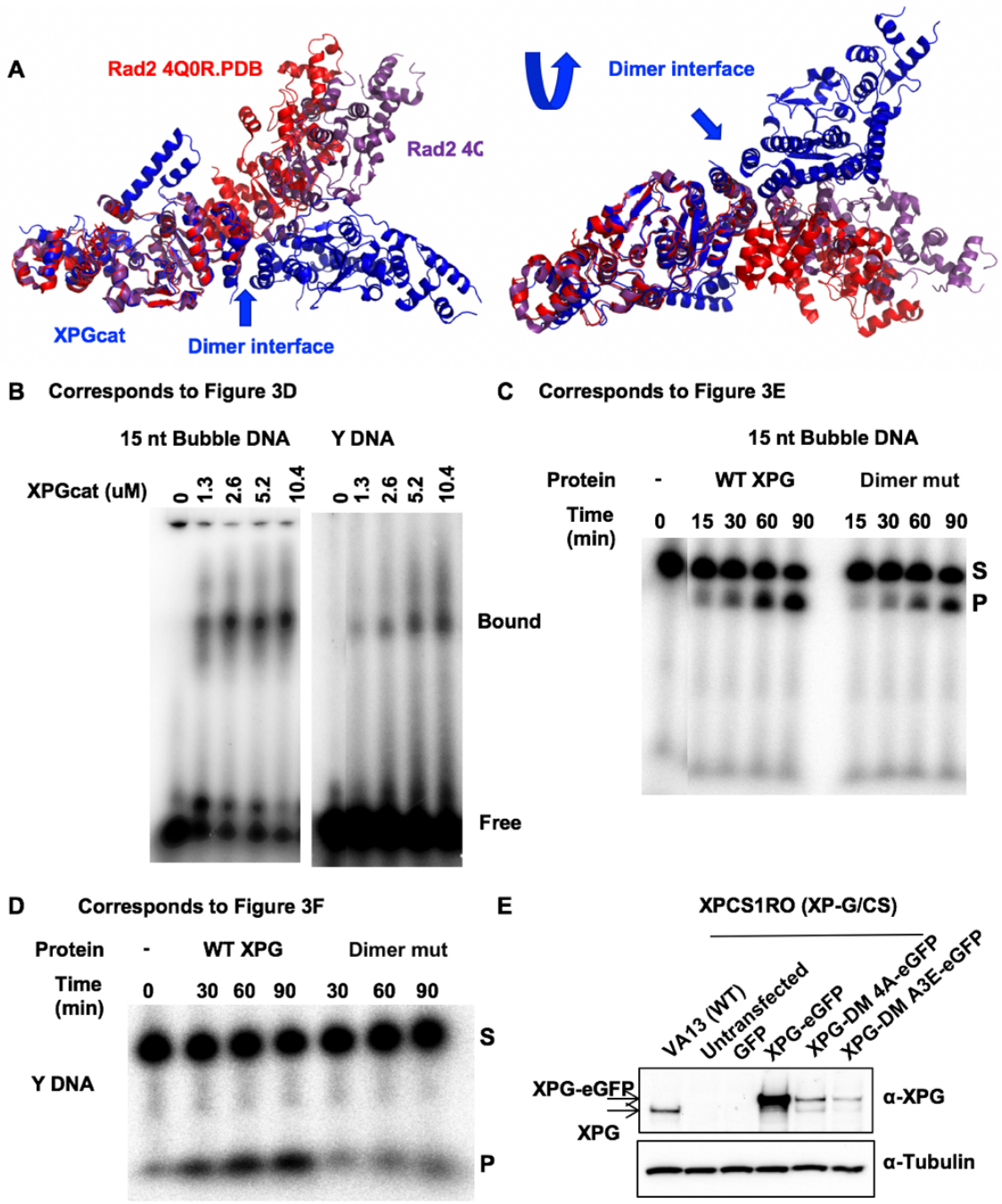


Fig. S4. Dimer model of XPG. A) Different perspectives of two Rad2 crystal forms of Rad2 overlaid on XPGcat crystal. Dimers based on the crystallographic lattice are depicted. The crystal contacts are similar but not identical between the yeast Rad2 and the human XPGcat. Different perspectives shown in the left and right reveal the details of the crystal contacts. **B**) Representative native gel of EMSA showing that XPGcat preferentially binds to 15 nt bubble DNA (HP15T/C) over Y DNA. Gel corresponds to Fig. 3D. **C and D**) Representative denaturing PAGE showing incision products of WT XPG and 4E Dimer mutant on 15 nt bubble DNA (HP15T/C) and Y DNA, as indicated. Gels correspond to Figs 3E-F, as indicated. **E)** Western analysis of XPG protein amounts in WT VA13 cells, SV40-transformed XP-G/CS cells from patient XPCS1RO transfected with an "empty" vector for eGFP expression only, or XPCS1RO cells transfected either with an expression vector for WT XPG-GFP or for dimer mutant XPG-GFPs (4A or A3E).

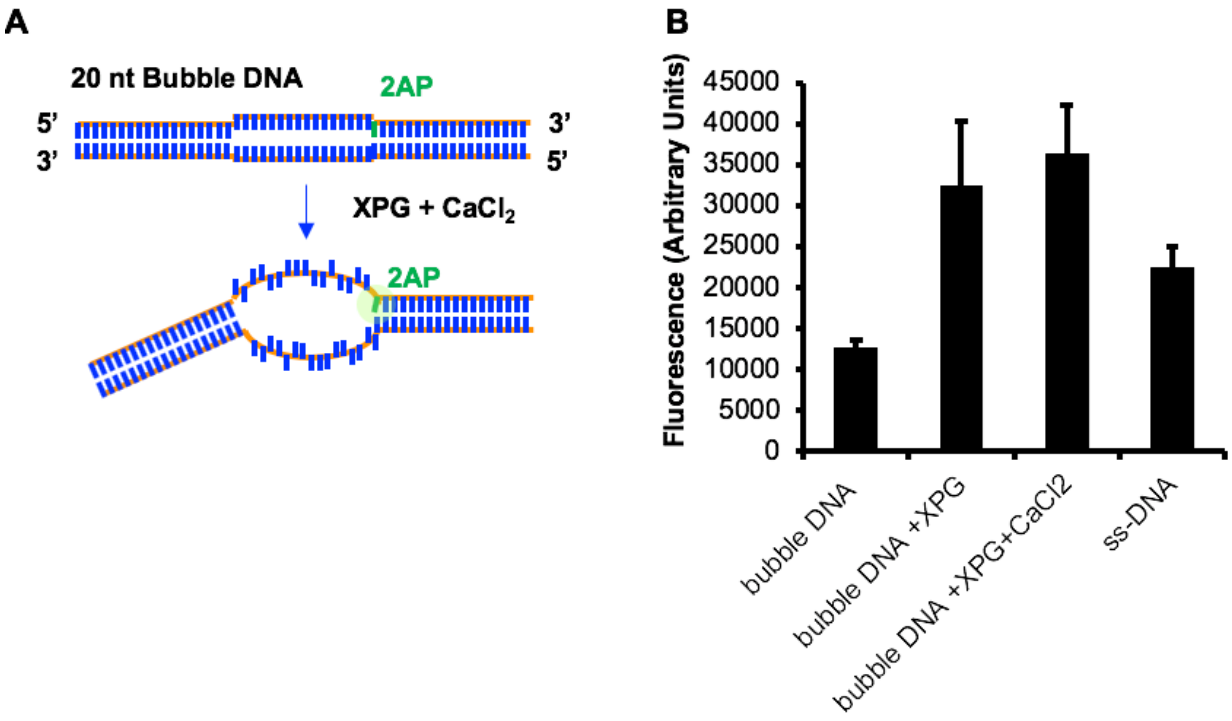


Fig. S5. Discontinuous dsDNA binding by XPG. **A**) Schematic of fluorescence substrates with one 2AP located at one ss/dsDNA junction of the DNA bubble. **B**) 2AP fluorescence experiment showing increased fluorescent signal in the presence of XPG, consistent with a break in stacking. Addition of $CaCl_2$, a divalent cation that mimics Mg^{2+} but is inactive, did not significantly alter the fluorescent signal.

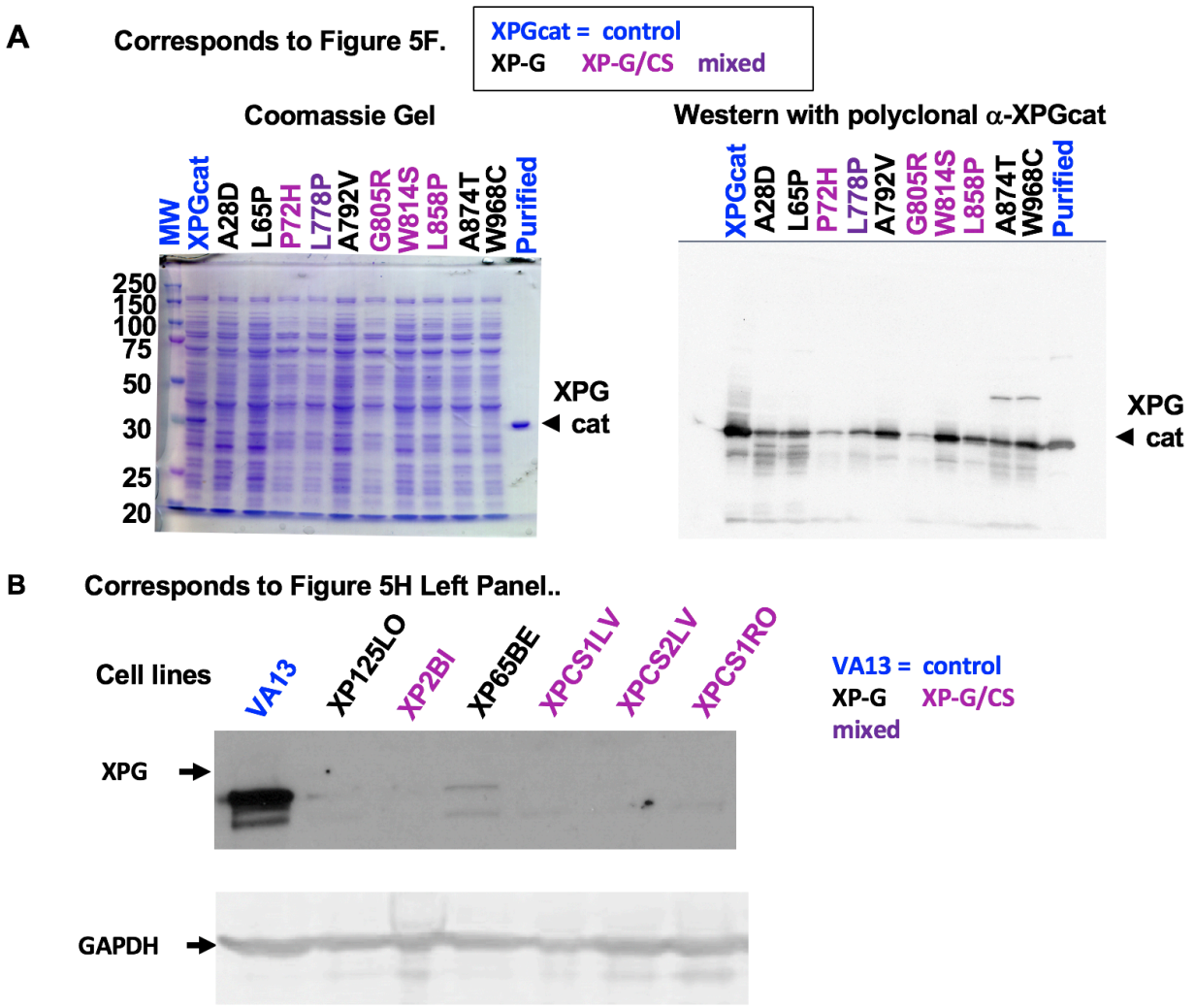


Fig. S6. ERCC5 Pathogenic Mutations. A) Disease mutations cloned into XPGcat did not express as well as the original XPGcat construct. Crude lysates of XP-G and XP-G/CS mutations made in XPGcat shown in Coomassie-stained SDS Page gel and Western corresponding to Figure 5F. Western probed with polyclonal antibody made against XPGcat. **B**) Shorter exposure for Figure 5H. Western analysis of both XP-G (black), XP-G/CS (plum), and mixed (purple) patient cell lines shows significant reduction in XPG protein levels, compared to SV40-transformed WT control VA13 fibroblasts or hTERT-immortalized WT control WI-38 fibroblasts, as appropriate. XPG detected with antibody R2; 97727 (1). The same membrane was re-probed with α-GAPDH antibody.

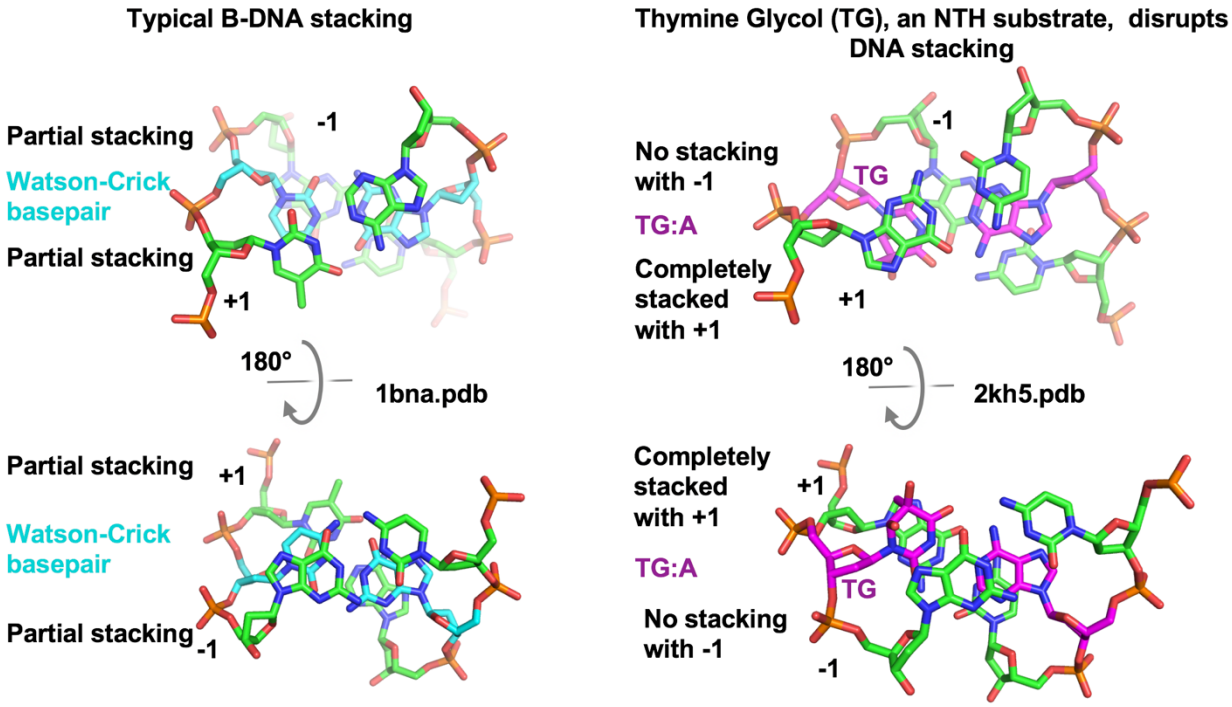


Fig. S7. Thymine glycol disrupts stacking in dsDNA. Top and bottom views of normal dsDNA and dsDNA containing a thymine glycol (TG). In B DNA, bases in each strand partially stack with each other. In DNA, the TG stacks fully with the base on the 5' side, but is not stacked with the base on the 3' side.

SUPPLEMENTARY TABLES.

Table S1. Data collection and refinement statistics.

Statistics for the highest-resolution shell are shown in parentheses.

	XPGcat
Wavelength	
Resolution range	43.35 - 2.0 (2.072 - 2.0)
Space group	C 2 2 21
Unit cell	64.4 173.4 101.6 90 90 90
Total reflections	481619
Unique reflections	35904 (2736)
Multiplicity	13.4(4.5)
Completeness (%) Scalepack	92 (66)
Completeness (%) PHENIX	83 (56.)
Mean I/sigma(I)	59 (1.8)
Wilson B-factor	43.5
R-meas	0.13 (0.80)
R-pim	0.03 (319)
Reflections used in refinement	32265 (2162)
Reflections used for R-free	1598 (106)
R-work	0.22 (0.33)
R-free	0.25 (0.38)
Number of non-hydrogen atoms	2889
macromolecules	2695
ligands	75
solvent	119
Protein residues	326
RMS(bonds)	0.012
RMS(angles)	1.25
Ramachandran favored (%)	97
Ramachandran allowed (%)	3
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	0.70
Clashscore	6.06
Average B-factor	72.8
macromolecules	71.8
ligands	113.
solvent	69.5

Table S2. DNA oligonucleotides used in this study.

Name	5'to3'sequence
5T-20 Y DNA	TTTTTGTGTCCTAGCACAGCGTATG
20-10C Y DNA	CATACGCTGTGCTAGGACACCCCCCCCCC
HP15T bubble	CGCGAAGCG GGGCAGACAACGTGGCGCTGTTTTTTTTTTTTTTTGTGTCCTAGCACAGCGTATG
HP15C bubble	CGCGAAGCG CATACGCTGTGCTAGGACACCCCCCCCCCCCCCCCAGCGCCACGTTGTCTGCCC
10T ds/ bubble	GGGCAGACAACGTGGCGCTGTTTTTTTTTTGTGTCCTAGCACAGCGTATG
10C bubble	CATACGCTGTGCTAGGACACCCCCCCCCCCCAGCGCCACGTTGTCTGCCC
10A ds	CATACGCTGTGCTAGGACACAAAAAAAAAACAGCGCCACGTTGTCTGCCC
SAXS10C	CAGCTCTGCCTCAGCCCCCCCCCCTCAAGACGGAACTA
SAXS10T	TAGTTCCGTCTTGATTTTTTTTTTCTGAGGCAGAGCTG
EM top	CGAATTCGAGTAAAACGACGGCCAGTGAAATTTGGTACCTGAGCAGTTCCCAGCTTGACTT CGTCCTCACTCTCTTCCTCTAGCGCTATATGCGTTGATGGACCAGAGGTGATAAGATGTCG ACGGAATTCTGAAGTAGGATTAATAGTAGTTTTTTTTTTCACAGAGAAGAACATTTGACCC GGGTAAAGCTAATAACAAGTAATCTGGTCCATCAACGCATATAGCGCTAGAGGAAGAGAGT GAGGACGAAGTCAAGCTGGGAACTGCTCAGGTACCAAATTTCACTGGCCGTCGTTT
EM bottom	AAACGACGGCCAGTGAAATTTGGTACCTGAGCAGTTCCCAGCTTGACTTCGTCCTCACTCT CTTCCTCTAGCGCTATATGCGTTGATGGACCAGATTACTTGTTATTAGCTTTACCCGGGTC AAATGTTCTTCTCTGTGCCCCCCCCCCCTACTATTAATCCTACTTCAGAATTCCGTCGACA TCTTATCACCTCTGGTCCATCAACGCATATAGCGCTAGAGGAAGAGAGTGAGGACGAAGTC AAGCTGGGAACTGCTCAGGTACCAAATTTCACTGGCCGTCGTTTTACTCGAATTCG
EMbubbletop	AGGTGATAAGATGTCGACGGAATTCTGAAGTAGGATTAATAGTAGTTTTTTTTTTCACAGA GAAGAACATTTGACCCGGGTAAAGCTAATAACAAGTAATCTGGTC
EMbubblebottom	ATTACTTGTTATTAGCTTTACCCGGGTCAAATGTTCTTCTCTGTGCCCCCCCCCCCTACTA TTAATCCTACTTCAGAATTCCGTCGACATCTTATCACCTCTGGTC
EM_armtop	CGAATTCGAGTAAAACGACGGCCAGTGAAATTTGGTACCTGAGCAGTTCCCAGCTTGACTT CGTCCTCACTCTCTTCCTCTAGCGCTATATGCGTTGATGGACCAG
EM_armbottom	CATCAACGCATATAGCGCTAGAGGAAGAGAGTGAGGACGAAGTCAAGCTGGGAACTGCTCA GGTACCAAATTTCACTGGCCpGTCGTTT
20T-AP bubble	GGGCAGACAACGTGGCGCTGTTTTTTTTTTTTTTTTTTT/i2AmPr/CGTCCTAGCACAG CGTATG
20C-bubble	CATACGCTGTGCTAGGACGTCCCCCCCCCCCCCCCCCCCCAGCGCCACGTTGTCTGCCC

Table S3. SAXS Data collection and metrics.

SAXS Data	XPG-flag	XPGΔC	XPGΔR	XPGΔR +DNA	XPGcat	XPGcat dimer mutant 4A
Porod Debye (Px)	2.9	3.2	3	3,2	3.9	3.9
Low q ($Å^{-1}$)	0.01308	0.01064	0.01064	0.0149	0.02098	0.01733
High q ($Å^{-1}$)	0.32805	0.32379	0.32379	0.32379	0.32255	0.32805
Reciprocal Rg (Å)	61	51	47	45	27	26
Reciprocal I(0 (detector units)	1100	27	270	170	170	750
Real space Rg (Å)	63	55	50	46	27	25
Real space I(0 (detector units)	1100	25	270	170	160	740
Dmax (Å)	210	178	167	159	106	99
Energy (keV)	12	11.111	11.111	11.111	12	12
Temp (°C)	21	20	20	20	10	10
Exposure (sec)	0.5	5	0.5	5	2	5
Protein Conc. (mg/ml)	0.52	0.196	1.62	mix	6.32	4.3

Table S4. Missense *xpg* Pathological Mutations with Structure-based Predictions. [1]Homozygous; [2]Heterozygous with 2 different missense mutations; [3]Heterozygous with stop codon or frameshift. [4] Human Gene Mutation Database and NIH patient cohort.

Mutant	Disease	Patient or database	Description (Reference(s))
G2W	XP-G	HGMD & NIH cohort[4]	Gly2, the N-terminus after Met1 processing, positions the catalytic water or metal ion. Mutation to tryptophan would distort metal ion and substrate position disrupting catalysis. (Pugh, 2019)
A28D	XP-G	XP01RJ[2] and XP02RJ[2]	Ala28 is in strand β2 buried in the hydrophobic core. Mutation to aspartate should disrupt a strand to helix connection under active site. Compound heterozygous with W968C. (Soltys, 2013)
L65P	XP-G	XP3HM[1]	Leu65 is in helix α3. Mutation to proline should disrupt the catalytic core at the ss/dsDNA transition. (Moriwaki, 2012; Sun, 2015)
P72H	XP-G/CS	XPCS4RO[3]	Pro72 in strand β3 forms a β–bulge and hydrophobic contact organizing the edge of the β–sheet. Histidine should alter core conformation and stability. (Zafeiriou, 2001)
I290N	XP-G/CS	XP101BR[1]	I290N is in a conserved region of the R-domain outside the XPGcat structure and thus highlighted orange, but mutation of hydrophobic β–branched isoleucine to asparagine may destabilize this region. (Fassihi, 2016).
L778P	XP-G XP-G/CS	XP40GO[3], XP174-1[1] XP174-2[1]	Leu778 at helix α7 C-terminus is between Arg777 and Lys957 and adjacent to a bound sulfate near the predicted 5' ssDNA. Mutation to proline should disrupt local packing of the C-terminal helices (α16 and α17) and DNA binding. XP40GO cells were biochemically similar to CS although no patient description exists, but XP174-1 and -2 were XP. This could potentially be explained by a dose effect, as XP40GO essentially had one XPG, while homozygous XP174-1 and -2 had two. (Schafer, 2013; Chikhaoui, 2019)
A792V	XP-G	XP125LO[3] XP124LO[3]	Ala792 on the helix α8 N-terminal end adjacent to conserved active site ligand 791. Mutation to valine should distort the active site. (Nouspikel and Clarkson,1994;Norris, 1987)
A795T	XP-G	XP118BR[3]	Ala795 one turn down on helix α8 from A792 such that a β–branched threonine side chain should shift 2 active site ligands. (Fassihi, 2016)
D798Y	XP-G/CS	HGMD and NIH cohort[4]	Asp798 on the C-terminal end of helix α8 H-bonds to His820 forming a junction of 3 structural elements. Mutation to tyrosine should disrupt core stability and the His820 β-hairpin for dsDNA binding. (Pugh, 2018)
G805R	XP-G/CS	XP165MA[1]	Gly805 starts strand β5. Arginine would destabilize the central β–sheet core by geometric collisions from the large side chain. (Schafer, 2013)
W814S	XP-G/CS	XP72MA[3]	Trp814 is a hydrophobic plug on helix α9 two-turn helix stabilizing 4 central structural elements. Serine should make a large cavity and destabilize the XPG core conformation. (Schafer, 2013)
A818V	XP-G	XP34BR[3], XP120BR[3], XP119BR[3]	Ala 818 is adjacent Trp814 and underlying metal ion ligands. The β–branched valine side chain should cause disruptive steric collisions impacting the active site. (Fassihi, 2016)
L858P	XP-G/CS Late onset	XP2BI[3]	Leu858 is a hydrophobic plug joining 3 structural elements at the C-terminal end of helix α11 in the H2TH near the K^+ site required for dsDNA binding in the FEN superfamily. Pro would disrupt this whole region. (Lalle, 2002)
A874T	XP-G	XP65BE[3]	Ala874 is on α12 helix of the H2TH packing with Leu858. The β–branched threonine side chain should disrupt the K^+ site and DNA binding. (Emmert, 2002)
W968C	XP-G	XP01RJ[2] and XP02RJ[2]	Trp968 on a loop joining helices α16-17 in a surface-exposed hydrophobic channel forming the proposed 3' ss DNA region. Mutation to cysteine should disrupt DNA binding. Compound heterozygous with A28D. (Soltys, 2013)