PNAS

Supplementary Information for

Two Interaction Surfaces between XPA and RPA Organize the Preincision Complex in Nucleotide Excision Repair

Mihyun Kim[1,2], Hyun-Suk Kim[1], Areetha D'Souza[3,4], Kaitlyn Gallagher[3,4], Eunwoo Jeong[1], Agnieszka Topolska-Wos[3,4], Kateryna Ogorodnik Le Meur[3,4], Chi-Lin Tsai[6], Miaw-Sheue Tsai[7], Minyong Kee[1], John A. Tainer[6], Jung-Eun Yeo[1], Walter J. Chazin[3,4,5,*], and Orlando D. Schärer[1,2,3,*]

[1]Center for Genomic Integrity, Institute for Basic Science, Ulsan, 44919 Republic of Korea; Department of Biological Sciences, [2]Ulsan National Institute of Science and Technology, Ulsan, 44919, Republic of Korea. [3]Department of Biochemistry, [4]Center for Structural Biology, and [5]Department of Chemistry, Vanderbilt University, Nashville, TN 37232-7917, USA. [6]Department of Molecular and Cellular Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA. [7]Biological and Systems Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA.

* Correspondence to Walter J. Chazin (walter.j.chazin@vanderbilt.edu) or Orlando D. Schärer (orlando.scharer@ibs.re.kr)

This PDF file includes:

p. 2-10 Supplementary Text: Extended Methods
p. 11 Supplementary Table 1
p. 12-17 Figures S1 to S6
p. 18 SI References

EXTENDED METHODS

Plasmids and mutagenesis- HA tagged XPA cDNA was cloned into a pWPXL vector. Mutated pWPXL-XPA and pBG100-XPA expression vectors were generated by site-directed mutagenesis from pWPXL-XPA WT and pBG100-XPA WT using the KOD-mutagenesis kit (Toyobo, TYB-SMK-101). The following primers were used for inverse PCR mutagenesis (altered codons underlined):
M70-1K-F: 5'-AAAGAATGTGGGAAAGAATTTATGG-3'
M70-1K-R: 5'-GCATATTACATAATCAAATTCCATAACAGGTCC-3'
M70-1A-F: 5'-GAAGCTATGGATTCTTATCTTATGAACCAC-3'
M70-1A-R: 5'-TTTCCCACATTCTTCGCATATTACA-3'
M32-1 (R34E): 5'-GAAGCACTGATGCTGCGCCAGGCC-3'
M32-1-R (R34E): 5'-CTGCCGCTTCCGCTCGATACTCG-3'
(XPA 70-M2_F primer is same with XPA 70-M1A_F primer.)
M70-2-R: 5'-TTTCCCACATTCTTTGCATATTACA-3'
M70-3-F (D101N/E106K/F112A): 5'-AATTATGTAATATGCAAAGAATGTGGGAA-3'
M70-3-R (D101N/E106K/F112A): 5'-AAATTCCATAACAGGTCCTGGTT-3'
M70-4-F (D101/E106K/K110E/F112A): 5'-GGGAAGAAGCTATGGATTCTTATC-3'
M70-4-F (D101/E106K/K110E/F112A): 5'-CACATTCTTTGCATATTACATAATTAAA-3'
M70-6-F (D101N/E106K/K110E/E111K/F112A/D114N): 5'-
AAAGCTATGAATTCTTATCTTATGAACC-3'
M70-6-F (D101N/E106K/K110E/E111K/F112A/D114N): 5'-TTCCCCACATTCTTTGCATATTAC-3' .
M32-3-F (R34E/R39E/R42E): 5'-GCCGAACTGGCTGCCCGGCCCTA-3'
M32-3-R (R34E/R39E/R42E): 5'-CTGTTCCAGCATCAGTGCTTCCT-3'
M32-4-F (R30E/K31E/R32E/R34E): 5'-GAAGAACAGGAAGCACTGATGCTGC-3'
M32-4-R (R30E/K31E/R32E/R34E): 5'-TTCCTCGATACTCGCCCGCACCGA-3'
(M32-6-F (R30E/K31E/R32E/R34E/R39E/R42E) primer is same with M32-4-F primer.
M32-6-R (R30E/K31E/R32E/R34E/R39E/R42E) primer is same with XPA M32-3-R primer).

XPA Protein expression and purification - Wild-type or mutant full-length XPA or truncated XPA_{1-239} with an N-terminal His_6 tag was expressed in *Escherichia coli* Rosetta pLysS cells using our previously described method (1). Cells were grown in TB medium containing 50 μg/ml kanamycin to OD_{600} = 0.6 at 37 °C, then OD_{600} = 1.2 at 18 °C, followed by induction with 0.5 mM isopropyl β-D-thiogalactoside at 18 °C overnight. Cells were collected by centrifugation at 6,500 rpm for 20 min and all subsequent purification steps were performed at 4 °C. The cell pellet from 1 L of culture was resuspended in 20 mL of lysis buffer (100 mM Tris-HCl at pH 8.0, 500 mM NaCl, 20 mM imidazole, 5 mM 2-mercaptoethanol, 10 μM $ZnCl_2$, 200 μg/ml lysozyme, 0.5 mM PMSF, 1 mM benzamidine, and 10% glycerol, and protease inhibitor (Roche)). The suspended pellets were Dounce homogenized 20 times and sonicated with an amplitude of 60 (5 sec on/10 sec off) for 2

min. The lysate was clarified by centrifugation at 20,000 rpm for 40 min and filtration of the supernatant through a 0.45 mm PVDF syringe filter. The supernatant was incubated for 90 min at 4 °C with 10 mL of Ni-resin pre-equilibrated with Ni-loading buffer (100 mM Tris-HCl at pH 8.0, 500 mM NaCl, 20 mM imidazole, 5 mM 2-mercaptoethanol, 10 μM $ZnCl_2$, and 10% glycerol). The resin was washed twice with Ni-wash buffer (100 mM Tris-HCl at pH 8.0, 500 mM NaCl, 20 mM imidazole, 5 mM 2-mercaptoethanol, 10 μM $ZnCl_2$, and 10% glycerol) and eluted with Ni-elution buffer (100 mM Tris-HCl at pH 8.0, 500 mM NaCl, 400 mM imidazole, 5 mM 2-mercaptoethanol, 10 μM $ZnCl_2$, and 10% glycerol). The eluent was incubated with H3C protease overnight at 4°C and dialyzed in buffer containing 20 mM Tris-HCl at pH 8.0, 300 mM NaCl, 5 mM 2-mercaptoethanol, 10 μM $ZnCl_2$, and 10% glycerol. The dialyzed sample was diluted to a concentration of 150 mM NaCl using a buffer of 20 mM Tris-HCl at pH 8.0, 5 mM 2-mercaptoethanol, 10 μM $ZnCl_2$, and 10% glycerol, then applied to a 5 mL Hi-trap heparin column equilibrated with Hep-buffer (20 mM Tris-HCl at pH 8.0, 50 mM NaCl, 5 mM 2-mercaptoethanol, 10 μM $ZnCl_2$, and 10% glycerol) and eluted using a linear gradient with Hep/NaCl-buffer (20 mM Tris-HCl at pH 8.0, 1 M NaCl, 5 mM 2-mercaptoethanol, 10 μM $ZnCl_2$, and 10% glycerol) over 5 column volumes. The XPA protein eluted at approximately 400 mM NaCl. The proteins were further purified on a HiLoad 16/600 Superdex 75pg column with buffer (50 mM Tris-HCl at pH 8.0, 150 mM NaCl, 1 mM dithiothreitol, 10 μM $ZnCl_2$, and 10% glycerol), where they eluted at 50-55 ml. The proteins were obtained at a concentration of ~1 mg/mL, yielding a total of 2-3 mg per liter of cell culture.

Isothermal Titration Calorimetry (ITC) - Protein samples were dialyzed against 20 mM Tris (pH 8.0), 150 mM NaCl, 3% glycerol, and 0.5 mM TCEP. Experiments were performed at 25 °C with 125 rpm stirring using an Affinity ITC instrument (TA Instruments). Titrations were performed by titrating an initial injection of 1 μL of 1.1 X 10^{-4} M XPA into the sample cell containing 0.20 X 10^{-4} M RPA, followed by an additional 47 injections of 3 μL each. The injections were spaced over 200-250 s intervals. Data were analyzed using NanoAnalyze software provided by TA Instruments. The thermodynamic parameters and binding affinities (K_d) were calculated using the average of three titrations fit to an independent binding model.

The M70-2 mutant is unstable at high concentrations and forms a precipitate during titrations. The precipitate caused mechanical issues during one of the replicates, and the last 10 injections were removed for analysis. To validate the K_d and stoichiometry calculations of the three forward titrations for the M70-2 mutant, we performed a reverse titration. The titration was conducted by injecting 1 μL of 90 X 10^{-6} M RPA into the sample cell containing 16 X 10^{-6} M M70-2 followed by an additional 47 injections of 3 μL each. Thermodynamic parameters similar to the forward titration were obtained.

Microscale Thermophoresis (MST) - DNA-binding affinity was measured using a Monolith NT.115 (NanoTemper Technologies). The XPA samples were buffer exchanged into 50 mM Tris-HCl (pH 7.8), 150 mM NaCl, 10 mM $MgCl_2$, 0.05% Tween-20, and 1 mM DTT and diluted into 16 concentrations ranging from 0.1 to 50 X 10^{-6} M. A 5'-6-FAM labelled 4-nucleotide hairpin substrate (5'–TTTTGCGGCCGCTTTTGCGGCCGC–3') was added to the protein at a final concentration of 40 X 10^{-9} M. All measurements were carried out at room temperature in standard capillaries (Nanotemper Technologies, Cat# MO-K022) using 20% excitation power and 20% MST power. K_d values were determined by fitting the data to the One Site – Total Binding model in GraphPad Prism version 8. Outlier data points were identified using a box plot analysis and removed if outside the whiskers.

Electrophoretic mobility shift assay (EMSA) - EMSA was conducted using three-way junction substrate as described in our previous paper (2). The annealed three-way junction oligonucleotides (100nM) with fluorescein-label were incubated with wild-type or mutant XPA (0, 20, 40, 80 nM) in a 10 µL mixture containing 25 mM Tris-HCl (pH 8.0), 1 mM DTT, 0.1 mg/ml BSA, 5% glycerol, and 1 mM EDTA at 25 °C for 30 min. The reaction mixture was loaded onto 8% native polyacrylamide gel and run at 4 °C for 2 hr at 20 mA with 0.5% TBE buffer. Gels were scanned using an Amersham Typhoon RGB imager. The band intensity was measured with ImageQuant TL program, and the fraction of bound XPA-bound protein quantified based on the intensity of the free DNA band.

Lentiviral Cell Line Transduction - SV40-transformed human fibroblasts XP2OS (XPA mutant) were cultured in Dulbecco's Modified Eagle's Medium (DMEM, Cytiva) supplemented with 10% fetal bovine serum (FBS, Millipore) and penicillin streptoCycin (P/S, Gibco) at 37°C in the presence of 5% CO2. For lentiviral transduction, 0.75 µg of pWPXL expression vector containing XPA cDNA, 2.25 µg of pMD2.G envelop plasmid, and 2.25 µg of psPAX2 packaging plasmid were transfected into 293T cells using Lipofectamine 3000 (L3000001, Thermo Fisher Scientific) following established protocols and virus harvested after 1 days (3). One day before transfection, XP2OS cells were seeded in 6-well plates at 50% confluency and incubated with lentivirus with a multiplicity of infection (MOI) of 2 for 24hr and subsequently grown as described above.

Western blot - Cells were lysed with RIPA buffer (150 nM NaCl, 50 mM Tris-HCl (pH 8.0), 5 mM EDTA, 1% Triton X-100, 0.1% SDS, 0.05 g/mL sodium deoxycholate, 1 mM phenylmethylsulfonyl fluoride (PMSF), 5 mM $MgCl_2$, supplemented with protease inhibitor (Roche)). Protein concentration was measured by Bradford assays (Thermo Fisher Scientific). AfterCaddition of LDS sample buffer containing 2.5% of 2-mercaptoethanol (4x, Invitrogen), the samples were heated to 95 °C for 10 min. Samples containing 25 µg of protein were separated in 8-16% Tris-Glycine gels (Thermo Fisher Scientific) at 150 V for 45 min, transferred to PVDF membrane (Amersham Hybond

0.2 um PVDF) at 100 V for 1hr 10 min with mini-protean tetra system (Bio-Rad) at 4 °C, and blocked with 5% skim milk in Tris-buffered saline containing 0.1% Tween-20 (TBST) for 1 hr at room temperature (RT). Rabbit polyclonal XPA antibodies (1:1,000), HA-tag antibodies (1:7,000), or Ku80 antibodies (1:3,000) were added to the TBST and incubated overnight at 4 °C. The blot was then incubated with anti-goat IgG rabbit antibody (1:3,000) for 1hr at RT and the bands visualized with chemiluminescent substrate (cat #: 34577, Thermo Fisher Scientific) and ChemiDoc touch imaging system (cat #: 1708370, Bio-Rad).

Local UV irradiation assay - Cells were seeded onto slide cover glass (Neuvitro, cat# GG-22-PDL) 1 day before UV treatment and incubated with growth medium. For UV treatment, the medium was removed, and cells were covered with 5 μm isopore membrane (cat #: TMTP04700, Merck) and irradiated with 100 J/m^2 of UV-C. After UV irradiation, cells were incubated for different repair times (0, 1, 2, 4, 8, 24 hrs for (6-4) PPs repair assay/ 0, 8, 24, 48 hrs for CPDs repair assay). After washing with DPBS, cells were fixed with 4% paraformaldehyde in DPBS for 10 min at RT and permeabilized with DPBS containing 0.5% Triton X-100 for 5 min on ice. For CPDs and (6-4) PPs detection, DNA was denatured with 2 M HCl for 30 min at RT. Following a 30 min blocking step with DPBS containing 20% FBS, cells were incubated with diluted primary antibody in 5% FBS for 2 hrs at RT. Following 5 washing steps with DPBS, the cells were incubated with diluted secondary antibody in 5% FBS for 1 hr at RT and stained with DAPI (Vectashield with DAPI, cat #: H-1200-10).

Clonogenic cell survival assay - Native and transduced XP2OS cells were cultured in DMEM supplemented with 10% FBS and 1% P/S. 2000 cells were seeded in a 6 cm dish 1 day before UV irradiation. Cells were treated with 0, 1, 2, 4 J/m^2 of UV-C and grown for 10 days. Cells were fixed with 4% paraformaldehyde for 15 minutes and stained with 1% of methylene blue for 2 hrs. After washing with water, colonies containing more than 25 cells were counted. The survival rate was normalized to the number of colonies of non-treated cells.

Co-localization assay - Cells were seeded and irradiated with UV-C as described for the local UV irradiation assay. After UV treatment, cells were lysed with cold hypotonic buffer (10 mM Tris-HCl (pH 8.0), 2.5 mM $MgCl_2$, 10mM β-glycerophosphate, 0.2% Igepal, 0.2 mM PMSF, 0.1 mM Na_3VO_4) for 15 min on ice and washed with hypotonic buffer without Igepal for 4 min at RT. Cells were fixed in 2% formaldehyde in DPBS for 5 min at RT. After washing twice with 0.1% Triton X-100 in DPBS for 5min, cells were stored in 70% ethanol at -20 °C. After removal of the ethanol, cells were washed with DNAseI digestion buffer (10mM Tris-HCl (pH 8.0), 5 mM $MgCl_2$) for 2 min and then the DNA digested with 20U DNaseI (cat #: D4527, Sigma-Aldrich) in 1 mL of digestion buffer for 40 sec at RT. EDTA (500 mM) was added to a final concentration is 10 mM EDTA to stop the reaction. The

solution was removed, and cells were washed twice with 10 mM EDTA in DPBS for 5min. After treatment with blocking buffer (1% BSA, 0.2% Tween-20), cells were incubated with diluted primary antibody (CPD (1:1,500) and HA-tag (1:250)) in blocking buffer for 2 hrs. Following 3 washing steps with DPBS containing 0.2% Tween-20, cells were incubated with diluted secondary antibody in blocking buffer and stained with DAPI.

Slot-blot assay - Cells were irradiated with 5 J/m^2 of UV-C and collected after different repair time (0, 2, 4, 8 hrs for (6-4) PPs and 0, 8, 24, 48 hrs for CPDs). Cells were harvested and genomic DNA isolated using the QIAamp DNA mini kit (Qiagen). Genomic DNA (200 ng of (6-4) PPs or 100 ng of CPDs) was denatured with 7.8 mM EDTA (for (6-4) PPs) or with 0.4 M NaOH and 10 mM EDTA (for CPDs), heated to 95 °C for 10 min, were neutralized by adding an equal volume of 2 M ammonium acetate (pH 7.0) and vacuum-transferred to a pre-washed nitrocellulose membrane using a BioDot SF microfiltration apparatus (Bio-Rad). Each well was washed 2 times with SSC buffer. The membrane was removed from the apparatus rinsed twice with SSC, air-dried, baked under vacuum at 80 °C for 2 hrs and blocked with 5% skim milk in PBS. For lesion detection, the membrane was incubated with (6-4) PP (1:2,000) or CPD antibodies (1:3,000) at 4 °C overnight and then incubated with anti-goat IgG mouse antibody (1:2,500 for (6-4) PPs or 1:5,000 for CPDs) for 1 hr at RT. The blot was visualized with ECL system (Thermo Fisher Scientific) and the total amount of DNA loaded on the membrane was visualized with SYBR-gold staining (Thermo Fisher Scientific).

TFIIH purification - TFIIH core complex (XPB-PreScission-GFP, XPD, p62, p52, p44, p34, and p8) was cloned into MacroBac vector 438A (4). The protein was expressed in Sf9 cells supplemented with 1 mM L-cysteine and 0.1 mM ferric ammonium citrate. Based on the XPB fusion with GFP, TFIIH was purified anaerobically using GFP-nanobody binder (5) covalently linked to agarose-beads (NHS agarose, Pierce/Thermo), eluted with PreScission protease, and purified using Superpose 6 (10/300) or Hi-Load Superdex 200 (16/600) for large-scale in 25 mM HEPES, pH 7.8, 150 mM NaCl, 50 mM KCl, 3% glycerol, and 3 mM 2-mercaptoethanol. The elution with TFIIH core complex (confirmed by SDS-PAGE gel and 420 nm absorption peak) was stored at -80 °C until use.

In vitro NER activity assay with cell extracts – A plasmid containing a site-specific acetylaminofluorene (AAF) lesion was incubated with XPA-deficient (XP2OS) cell extract in the absence or presence of purified wild-type or mutant XPA proteins as described previously (6). For each reaction, 2 µl of repair buffer (200 mM HEPES-KOH, 25 mM $MgCl_2$, 110 mM phosphocreatine (di-Tris salt, Sigma), 10 mM ATP, 2.5 mM DTT and 1.8 mg/ml BSA, adjusted to pH 7.8), 0.2 µl of creatine phosphokinase (2.5 mg/ml, Sigma), 3 µl of XPA-deficient cell extract (about 10 mg/ml),

NaCl (to a final concentration of 70 mM) and 50 nM of purified wild-type or mutant XPA in a total volume of 9 µl were pre-warmed at 30 °C for 10 min. 1 µl of plasmid containing AAF (25 ng/µl) was added to reaction mixture and incubated at 30 °C for different incubation times (0, 5, 10, 20, 45, and 90 min). 0.5 µl of 1 µM of a 3′-phophorylated oligonucleotide for product labeling was added and the mixture heated at 95 °C for 5 min. The mixture was allowed to cool down to room temperature for 15 min. 1.2 µl of a Sequenase/[α-^{32}P-dCTP mix (0.25 units of Sequenase and 2.5 µCi of [α-^{32}P]-dCTP per reaction) was added and the mixture incubated at 37 °C for 3 min. Then 1.2 µl of dNTP mix (100 µM of each dATP, dTTP, dGTP; 50 µM dCTP) added to mixture and incubated for another 12 min. The reactions were stopped by adding 12 µl of loading dye (80% formamide/10 mM EDTA) and heating at 95 °C for 5 min. 6 ul of sample was loaded on 14% sequencing gel (7 M urea, 0.5x TBE) at 45 W for 2 ½ hrs. The reaction products were visualized using a PhosphorImager (Amersham Typhoon RGB, GE Healthcare Bio-Sciences). Two independent repetitions were performed. The NER products were quantified using ImageQuant TL and normalized to the amount of NER product formed with WT-XPA at 90 min.

Small-angle X-ray scattering (SAXS) - The SAXS data were collected in SEC-SAXS mode at the ALS beamline 12.3.1 LBNL Berkeley, California (7). The X-ray wavelength λ was 1.03 Å and the sample-to-detector distance was set to 1.5 m resulting in scattering vectors, q, ranging from 0.01 Å^{-1} to 0.5 Å^{-1}. The scattering vector is defined as $q = 4\pi \sin\theta/\lambda$, where 2θ is the scattering angle. All experiments were performed at 20 °C (8) and data were processed as described (9). Briefly, the flow through SAXS cell was directly coupled with an online Agilent 1260 Infinity HPLC system using a Shodex KW803 column (Shodex^{TM}). The column was equilibrated with running buffer (20 mM Tris pH 8.0, 150 mM NaCl, 2% glycerol, 1 mM DTT) with a flow rate of 0.5 mL/min. A 50 µL sample of the complex (prepared in 1:1:1 stoichiometry with a final concentration of 25µM) was run through the SEC and 3.0 second X-ray exposures were collected continuously during a ~35 min elution. The SAXS frames recorded prior to the protein elution peak were used as buffer blanks to subtract from all other frames. The subtracted frames were examined by radius of gyration (R_g) and scattering intensity at q = 0Å^{-1} (I(0)), derived using the Guinier approximation $I(q) = I(0) \exp(-q2R_g2/3)$ with the limits $qR_g < 1.3$). I(0) and R_g values were compared for each collected SAXS curve (frame) across the entire elution peak (Fig. S4A). The elution peak was mapped by plotting the scattering intensity at q=0 Å^{-1} (I(0)), relative to the recorded frame. Uniform R_g values across an elution peak represented a homogenous assembly (Fig. S4A). The merged experimental SAXS data were additionally investigated for aggregation by inspecting Guinier plots (Fig. S4*B*). The program SCÅTTER 4.0 was used to compute the distance distribution, P(r) (Fig. S4*E*). The distance r where P(r) approaches zero intensity is termed the D_{max} of the molecule. P(r) was normalized based on molecular mass (142kDa for the XPA-RPA-DNA FL complex and 131kDa for XPA$_{1-239}$-RPAΔ32NΔ70N-DNA complex) determined by SAXS as calculated by SCÅTTER as described (10).

The differences in the scattering power of protein and DNA were not taken into account in the determination of molecular mass because the contribution of DNA to the total mass is much less than that of protein. The SAXS data for the complexes are deposited in the SASBDB databank with the accession codes: SASDPZ3 and SASDP24 for the 3' and 5 junction substrates, respectively.

SAXS data analysis - High quality SAXS data were collected for the complex of XPA and RPA with the 3' junction substrate. Analysis of the data revealed a monodisperse complex with a substantial degree of residual flexibility evident in the Kratky Plot and a value of 3.0 for the Porod exponent (**Table S1**). Importantly, residual flexibility to this extent greatly limits structural interpretation of the data (11). Both XPA and RPA have substantial flexible regions outside the two points of contact. For XPA this includes the C-terminal domain ($XPA_{240-273}$) and for RPA the disordered RPA32N domain ($RPA32_{1-42}$) and the globular RPA70N connected to RPA70A by a 68-residue flexible linker. We surmised that these regions are the origin of the residual flexibility observed in the SAXS data for the full-length proteins. Since these flexible regions do not contribute to the interactions between XPA and RPA or to binding the DNA substrate, we prepared the corresponding truncation constructs for further SAXS analysis: XPA_{1-239} and RPAΔ32NΔ70N (including deletion of the long linker to RPA70N). As for the full-length proteins, mixing of truncated variants of XPA and RPA with the same DNA substrate provided a very stable complex.

SAXS data were collected first for the complex of XPA_{1-239} and RPAΔ32NΔ70N bound to the 3' junction substrate. The SAXS profile and linearity of the Guinier plot (**Fig. S5B**) showed that the complex was free of aggregation in solution. The Kratky plot revealed both complexes contain globular domains and some flexible loops and/or linkers (**Fig. S5D**). Porod-Debye analysis, including the value of 3.5 for the Porod Exponent, indicated the ternary complex is stable and globular, and suggested that direct analysis of the distance distribution function, P(r), in terms of a molecular shape is feasible (**Table S1, Fig. S5F**). The primary peak centered at ~50 Å in the P(r) is attributed to distance distributions within the globular domains. The small feature in P(r) at ~120 Å is consistent with distances between regions of XPA and RPA that do not contact each other. SAXS data collected for XPA_{1-239}/RPAΔ32NΔ70N bound to the 5' junction substrate were of similar high quality, consistent with a well-formed and globular structure as reflected in the Porod Exponent of 3.4. The secondary shoulder in the P(r) is more prominent in the 5' junction complex at ~150 Å, suggesting there may be more flexibility within this complex. *Ab initio* shape calculations performed with DENSS provided molecular envelopes for the two complexes (**Fig. 6, S6**).

Previously determined SAXS analyses of the XPA-DBD/RPA70AB-DNA complex and of the RPA DNA binding apparatus (RPA-DBC) bound to ssDNA were used to interpret the molecular envelope for the 3'junction substrate complex (**Fig. S6**) (2,12). The SAXS based structures of RPA-DBC bound to a 30-nt ssDNA oligomer and of XPA-DBD/RPA70AB bound to a Y-shaped ss-ds DNA junction were superposed by aligning the RPA70AB domains. Next, a homology model of the

RPA32C bound to XPA_{29-46} (13) was generated using Modeller (14). The RPA32C/XPA_{29-46} complex was then localized within the molecular envelope of the full complex using FoXSDock (11, 12). The representative structure selected for further analysis had the best combined score of χ fit to the scattering curve and energy. Disordered residues were added at the XPA N-terminus (1-28), between XPA_{29-46} and XPA-DBD (47-97), and the linker residues between RPA32D and RPA32C ($RPA32_{172-199}$) using Modeller (15).The final representative structural model shows a reasonable fit to the experimental scattering profile (χ= 2.8) and molecular envelope of the complex. This representative model exhibits a curved and compact arrangement of domains. The trimer core (RPA70C/32D/14) and the RPA32C/XPA_{29-46} are positioned within the larger lobe of the molecular envelope, and XPA-DBD/RPA70AB in the smaller extended lobe (**Fig. 6A**).

The same overall approach was used interpreting the molecular envelope for the 5' junction substrate complex (**Fig. S6**). In this case, we were unable to achieve a good fit to the experimental data using the structures of the XPA-DBD/RPA70AB/DNA and RPA-DBC/DNA as was case for the 3' junction substrate complex. Rather, the best fit to data was achieved by docking the structures of the XPA-DBD/RPA70AB/DNA complex and the RPA trimer core (RPA70C/32D/14) with ssDNA bound extracted from SAXS-MD structure RPADBC bound to 20-nt ssDNA using FoXS Dock (32-33,43-44). The best representative scoring model was then further docked with the RPA32C/XPA_{29-46} homology model. As for the 3'junction substrate complex, disordered residues were added using Modeller (15,16). This included at the N-terminus of XPA (1-28), between XPA_{29-46} and XPA-DBD (47-97), the linker between RPA32D and RPA32C ($RPA32_{172-199}$), the linker between the RPA70B and RPA70C domains ($RPA70_{423-435}$), and the additional ssDNA (10 nucleotides) spanning from the 3' end of the ssDNA bound to RPA70B to the 5' end of the ssDNA bound to RPA 32C. The final representative structural model shows a reasonable fit to the experimental scattering profile (χ= 4.75) and molecular envelope of the complex. This representative model exhibits a similar curved and compact arrangement of domains with the trimer core (RPA70C/32D/14) and RPA32C/XPA_{29-46} positioned within the larger lobe of the molecular envelope, and XPA-DBD/RPA70AB in the smaller extended lobe (**Fig. 6B**).

Computational Modeling - The structure of the XPA_{1-239}/DNA/RPAΔ32NΔ70N complex was generated in four steps:

1. Two different initial models of the XPA_{98-239}/RPA DNA binding core (RPA70AB + RPA70C/32D/14, RPA-DBC) complex bound to a model DNA substrate were prepared by using previously determined integrative structures (2, 12). The substrate consisted of a Y-shaped DNA with a 10-bp duplex, a 8-nt 5' overhang and a 30-nt 3'overhang. Model A was generated by combining (i) a jointly refined SAXS and molecular dynamics simulation structure of RPA DBC bound to a 30-nt ssDNA substrate (10) and (ii) a model of XPA_{98-239}/RPA70AB complex bound to a 3' ss-ds DNA junction generated from a combination of homology modeling, NMR, SAXS, and

molecular docking (PDBDEV 0039) (2). RPA70AB is contained in both, so the structures could be aligned by superposing them on these two domains using PyMOL. Model B was generated using FoXS Dock by combining the same integrative model of the XPA_{98-239}/RPA70AB/DNA complex with the RPA trimer core (RPA70C/RPA32D/RPA14) without any pre-alignment.

2. A homology model of RPA32C bound to XPA_{29-46} was prepared using the NMR structure of RPA32C bound to XPA homologue UNG2 (13, PDB: 1dpu) as a template in Modeller 9v4 (14).

3. FoXS Dock was used to dock the structures from steps 1 and 2 using the SAXS profile of the full XPA_{1-239}/DNA/RPAΔ32NΔ70N complex (15,16). A total of 4878 models were generated and the top model selected had the best combined SAXS χ and statistical potential scores. Although the results obtained were effectively the same for the two different initial models generated in step 1, the structural model for the complex with the Y-shaped DNA substrate composed of a 10-bp duplex, a 30-nt 3' overhang and a 8-nt 5'overhang used for further refinement were derived from initial model A.

4. Missing residues in XPA (N-terminal residues 1-28 and residues 47-97) and the linker between RPA32D and RPA32C were built using the program Modeller 9v4 (14,17). Then the FoXS server was used to best fit the model to the experimental SAXS profile.

The same strategy was used to generate the structural model of the complex with the Y-shaped DNA substrate composed of a 10-bp duplex, a 30-nt 5' overhang and a 8-nt 3'overhang. In this case, in step 3 a better fit to the data was obtained for structural models derived from initial model B. The integrative structural models will be deposited in the PDB-DEV databank.

Visualization and superposition of models into the molecular envelopes generated by DENSS (18) were made using the Chimera Fit to Map tool (19) and DENSS align (18).

Supplementary Table 1. Summary of parameters obtained by SEC-SAXS for XPA-RPA-DNA complexes

Constructs	D_{max} (Å)	R_g (Real) / R_g (Reciprocal) (Å)	Porod Exponent
XPA_{1-239}/3' junction/RPAΔ32NΔ70N	146.5	43.25/ 42.74	3.5
XPA_{1-239}/5' junction/RPAΔ32NΔ70N	164.5	45.83/ 46.81	3.4
XPA FL/3' junction/RPA FL	196	51.8/ 49.75	3.0

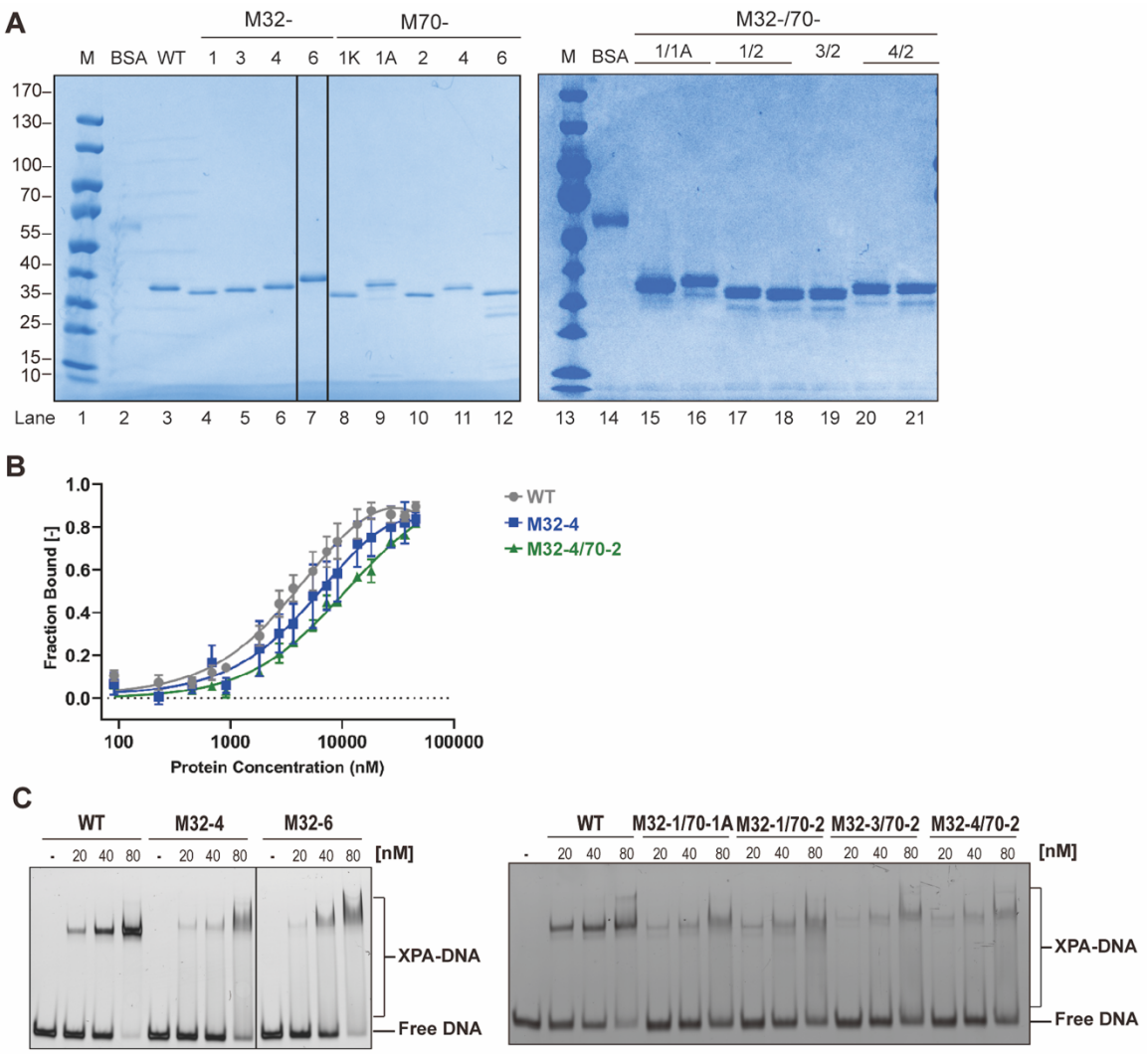

Fig. S1. DNA binding affinity of XPA wild-type, XPA-32-4E, and XPA-34E470-M2. A. Purification of XPA-RPA mutant proteins. **B.** The mean and standard deviation of 3 independent MST measurements of K_d values for the DNA substrate by wild-type, 32C-4E, and 32C-4E/70-M2 XPA were 5.4 ± 1.1 μM, 8.4 ± 2.3 μM, and 10 ± 3 μM, respectively. Error bars indicate the standard deviation (*N* = 3). **C.** EMSA analysis of XPA proteins with mutation in the RPA32 and RPA30/70 interaction domains binding to a 5'-FAM-labeled three-way DNA junction (100 nM). Protein concentrations and position of DNA on the gel are indicated.

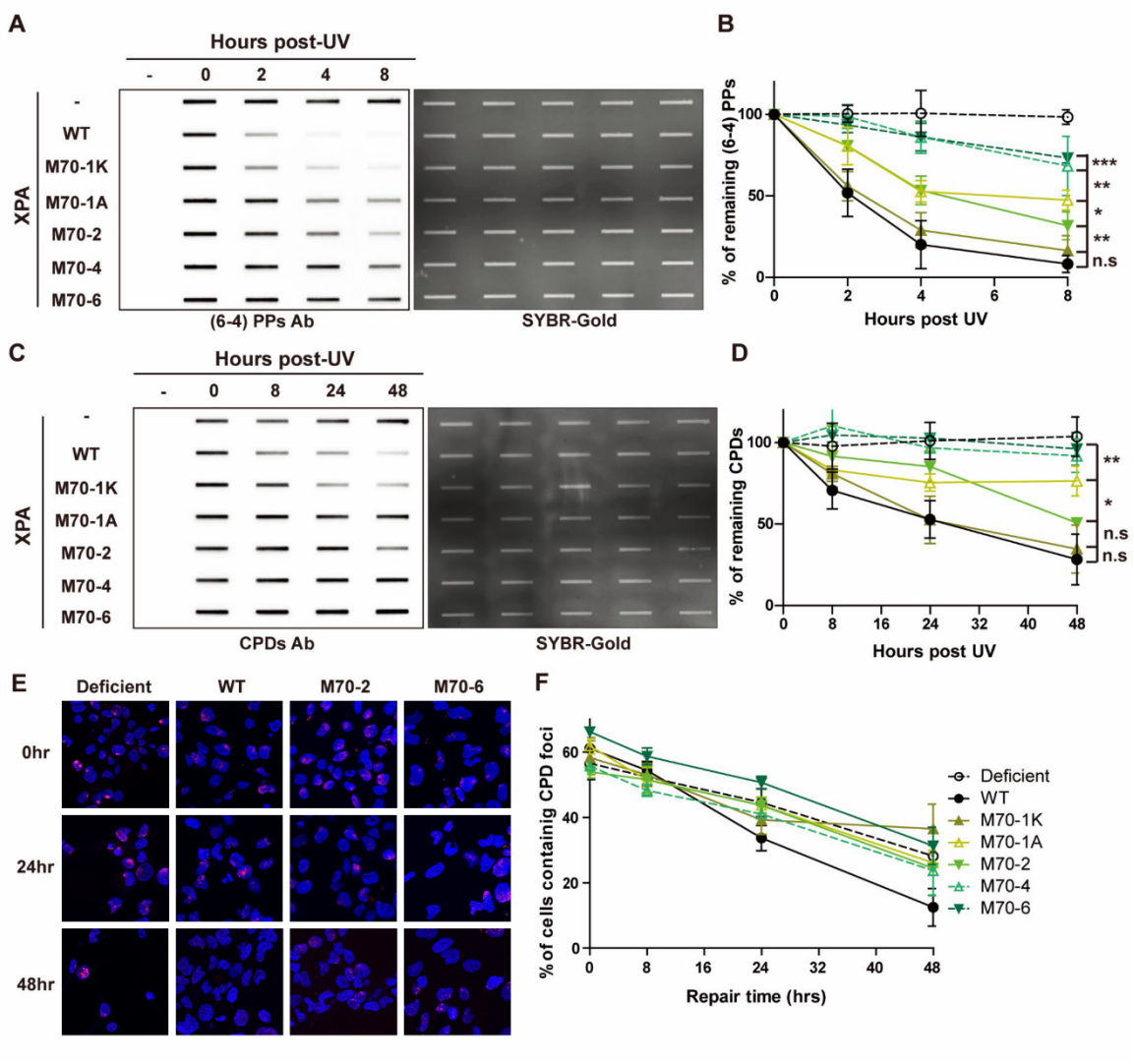

Fig. S2. (6-4) PP and CPD repair kinetics of XPA-RPA70 interaction mutants of XPA. A. Determination of (6-4) PPs repair kinetics using slot-blot assays. Cells were irradiated with 5 J/m^2, genomic DNA isolated at indicated time point and the adduct levels determined using an anti-(6-4) PPs antibody. The right panel shows total DNA stained with EtBr. **B.** Quantification of A. Band intensity was normalized to the WT band at 0 hr. The data represent 3 independent experiments. The p-value was measured compared to XPA WT. *P< 0.05, **P<0.01, ***P<0.001. **C.** Determination of CPD repair kinetics. Slot blots were conducted as in A, except that an anti-CPD antibody was used. **D.** Quantification of C. Band intensity was normalized to the WT band at 0 hr and the data represent 3 independent experiments. The p-value was measured compared to XPA WT. *P< 0.05, **P<0.01, ***P<0.001. **E.** Representative figures of cells irradiated through a 5 μm micropore filter with UV irradiation 100J/m^2 and stained for CPD to measure repair kinetics. CPD foci are red and cell nuclei are stained blue with DAPI. **F.** Quantification of E. 100 cells were counted for each sample and the data represent at least 2 independent experiments.

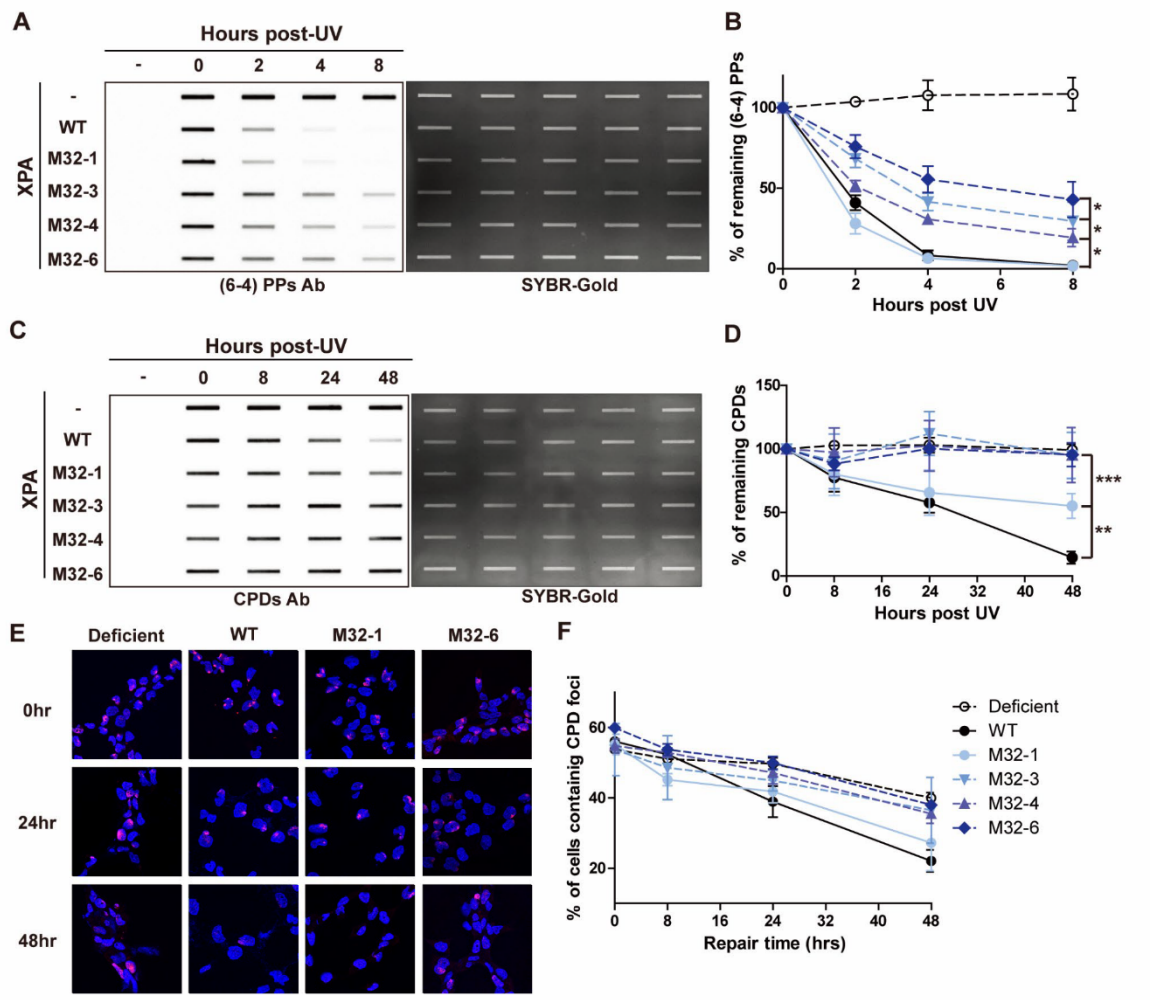

Fig. S3. (6-4) PPs and CPDs repair kinetics of RPA32 interaction mutants of XPA. A. Determination of (6-4) PPs repair kinetics using slot-blot assays. Cells were irradiated with 5J/m^2 and genomic DNA isolated at indicated time point and the adduct levels determined using an anti-(6-4) PP antibody. The right panel shows total DNA stained with EtBr. **B.** Quantification of A. Band intensity was normalized to the WT band at 0 hr. The data represent 3 independent experiments. The p-value was measured compared to XPA WT. *P< 0.05, **P<0.01, ***P<0.001. **C.** Determination of CPD repair kinetics using slot-blot assays. Slot blots were conducted as in A, except that an anti-CPD antibody was used. **D.** Quantification of C. The data represent 3 independent experiments. The p-value was measured compared to XPA WT. *P< 0.05, **P<0.01, ***P<0.001. **E.** Representative figures of cells irradiated through a 5 μm micropore filter with UV (100J/m^2) and stained for CPD to measure repair kinetics. CPD foci are red and nuclei are stained blue with DAPI. **F.** Quantification of (E). 100 cells were counted for each sample and the data represent at least 2 independent experiments.

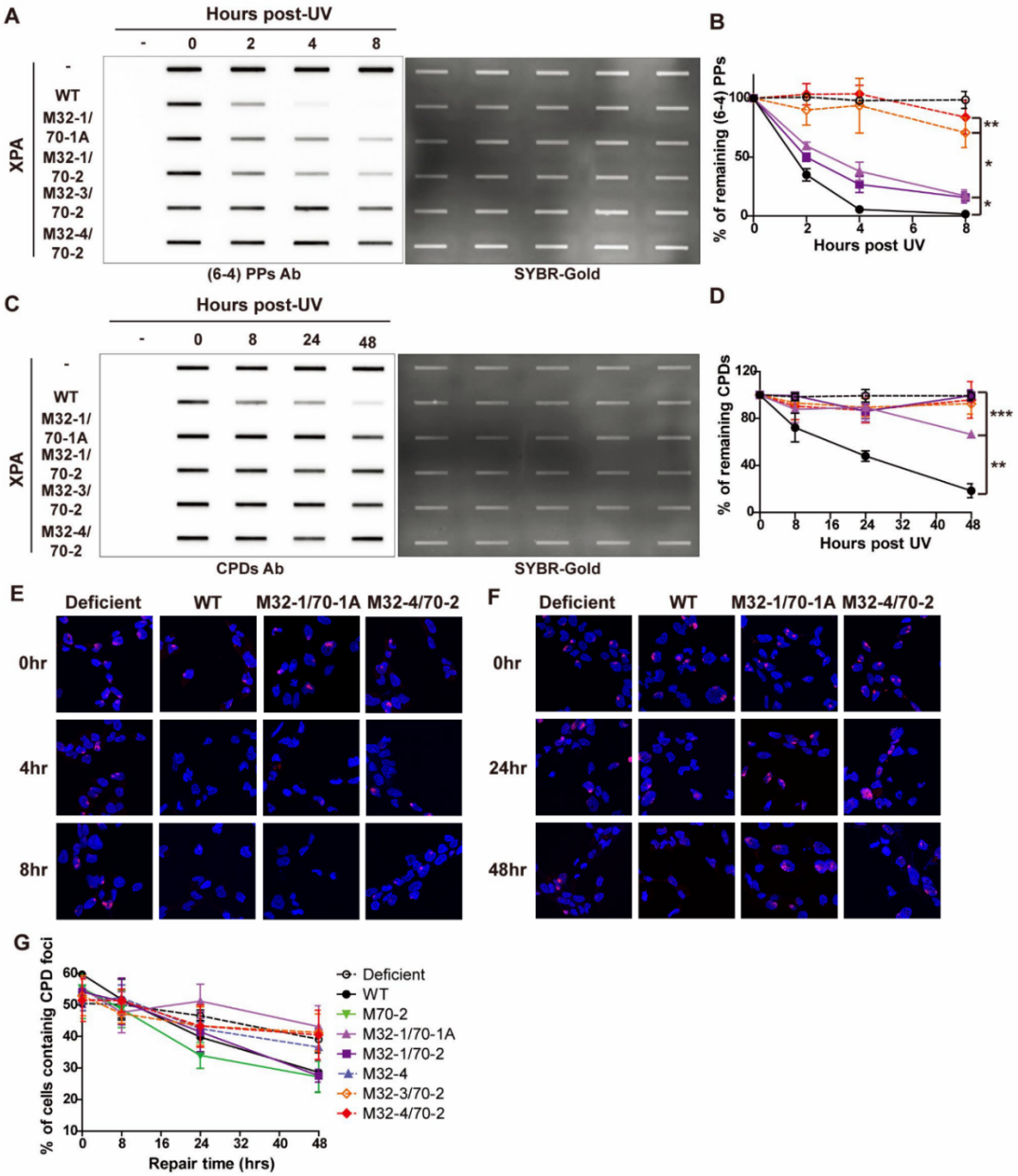

Fig. S4: (6-4) PPs and CPDs repair kinetics of RPA32/70 interaction mutants of XPA. A. Determination of (6-4) PPs repair kinetics using slot-blot assays. Cells were irradiated with 5 J/m^2 and genomic DNA isolated at indicated time point and the adduct levels determined using an anti-(6-4) PP antibody. **B.** Quantification of A. Band intensity was normalized to the WT band at 0 hr and the data represent 3 independent experiments. The p-value was measured compared to XPA WT. *P< 0.05, **P<0.01, ***P<0.001. **C.** Determination of CPD repair kinetics using slot-blot assays. Slot blots were conducted as in A. except that an anti-CPD antibody was used. **D.** Quantification of C. The data represent 3 independent experiments. The p-value was measured compared to XPA WT. *P< 0.05, **P<0.01, ***P<0.001. **E.** Representative figures of cells irradiated through a 5 μm micropore filter with UV (100J/m^2) and stained for (6-4) PPs to measure repair kinetics. (6-4) PP foci are red and the cell nuclei are stained blue with DAPI. **F.** Representative figures of cells irradiated through a 5 μm micropore filter with UV (100J/m^2) and stained for CPD to measure repair kinetics. CPD foci are red and the cell nuclei are stained blue with DAPI. **G.** Quantification of F. 100 cells were counted for each sample and the data represent at least 2 independent experiments.

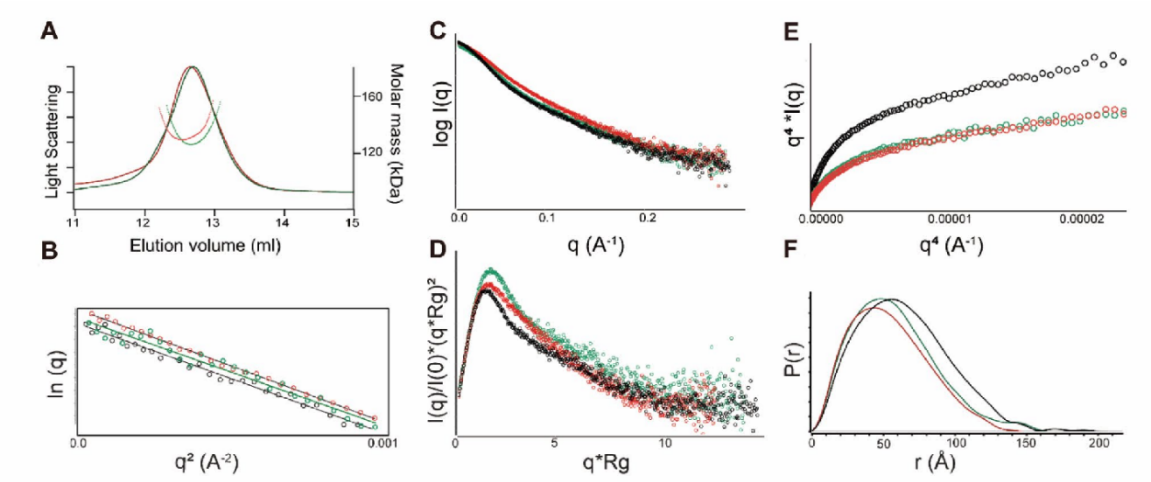

Fig S5. SEC-SAXS analysis of XPA FL/3' junction/RPA FL complex (black), XPA$_{1-239}$/3'junction/RPAΔ32NΔ70N complex (red) and XPA1-239/5'junction/RPAΔ32NΔ70N complex (green). **A.** SEC-MALS trace showing the molar mass values (dots) observed across the elution profile. **B.** Guinier plots for the two complexes show that samples are free of aggregation. **C.** The full scattering profile for the three complexes. **D.** Kratky plots **E.** Porod plots. The Porod Exponent was estimated as 3.0 for the FL complex and 3.5 for the truncated complexes. **F.** Distance distribution analysis for the three complexes.

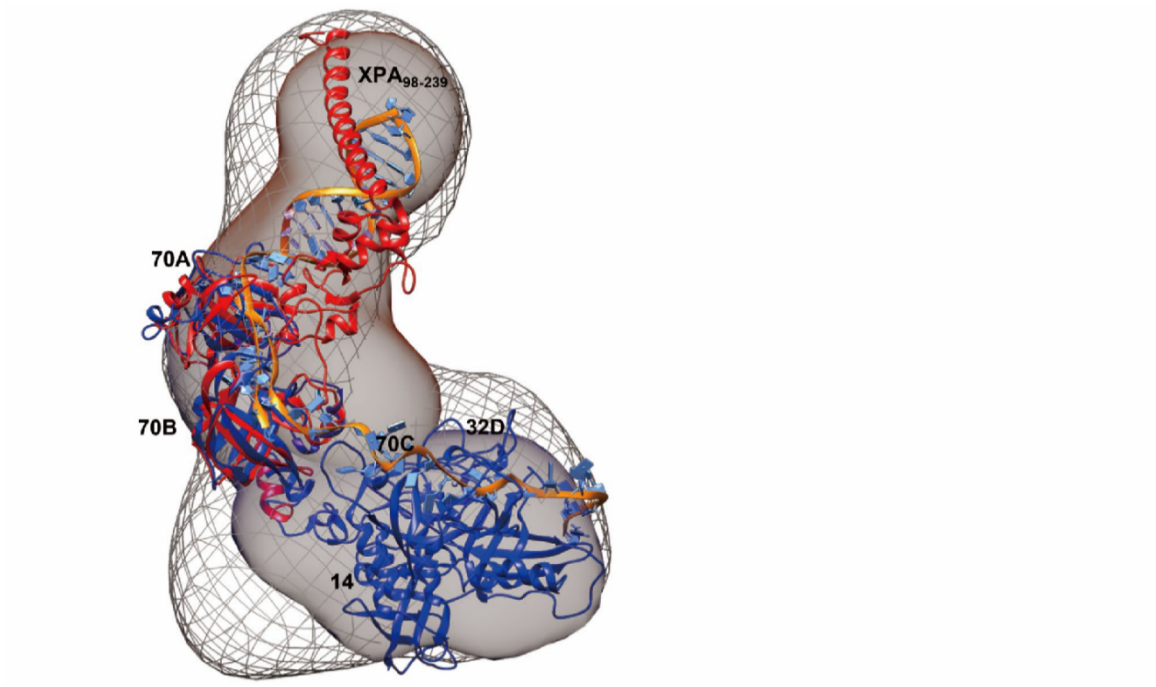

Figure S6: Molecular envelope of XPA_{1-239}/3'junction/RPAΔ32NΔ70N complex (mesh) compared with SAXS molecular envelopes of RPA DNA binding core (blue) bound to 30nt ssDNA and XPA_{98-239}/DNA/RPAΔ32NΔ70N complex (red) prepared using DENSS. The structures are superimposed at RP70AB domains. The unfilled region in the molecular envelope would correspond to RPA 32C/XPA_{29-46} complex.

SI REFERENCES

1. N. Sugitani, S. M. Shell, S. E. Soss, W. J. Chazin, Redefining the DNA-binding domain of human XPA. *J Am Chem Soc* **136**, 10830-10833 (2014).
2. A. M. Topolska-Woś *et al.*, A key interaction with RPA orients XPA in NER complexes. *Nucleic Acids Res* **48**, 2173-2188 (2020).
3. P. Salmon, D. Trono, Production and titration of lentiviral vectors. *Curr Protoc Neurosci* **Chapter 4**, Unit 4 21 (2006).
4. S. D. Gradia *et al.*, MacroBac: New Technologies for Robust and Efficient Large-Scale Production of Recombinant Multiprotein Complexes. *Methods Enzymol* **592**, 1-26 (2017).
5. A. Kirchhofer *et al.*, Modulation of protein properties in living cells using nanobodies. *Nat Struct Mol Biol* **17**, 133-138 (2010).
6. O. V. Tsodikov *et al.*, Structural basis for the recruitment of ERCC1-XPF to nucleotide excision repair complexes by XPA. *EMBO J* **26**, 4768-4776 (2007).
7. S. Classen *et al.*, Implementation and performance of SIBYLS: a dual endstation small-angle X-ray scattering and macromolecular crystallography beamline at the Advanced Light Source. *J Appl Crystallogr* **46**, 1-13 (2013).
8. K. N. Dyer *et al.*, High-throughput SAXS for the characterization of biomolecules in solution: a practical approach. *Methods Mol Biol* **1091**, 245-258 (2014).
9. G. L. Hura *et al.*, Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). *Nat Methods* **6**, 606-612 (2009).
10. R. P. Rambo, J. A. Tainer, Accurate assessment of mass, models and resolution by small-angle scattering. *Nature* **496**, 477-481 (2013).
11. R. P. Rambo, J. A. Tainer, Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers* **95**, 559-571 (2011).
12 C. A. Brosey *et al.*, A new structural framework for integrating replication protein A into DNA processing machinery. *Nucleic Acids Res* **41**, 2313-2327 (2013)
13. G. Mer *et al.*, Structural basis for the recognition of DNA repair proteins UNG2, XPA, and RAD52 by replication factor RPA. *Cell* **103**, 449-456 (2000).
14. B. Webb, A. Sali, Comparative Protein Structure Modeling Using MODELLER. *Curr Protoc Protein Sci* **86**, 2 9 1-2 9 37 (2016).
15. D. Schneidman-Duhovny, M. Hammel, J. A. Tainer, A. Sali, FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. *Nucleic Acids Res* **44**, W424-429 (2016).
16. D. Schneidman-Duhovny, M. Hammel, A. Sali, Macromolecular docking restrained by a small angle X-ray scattering profile. *J Struct Biol* **173**, 461-471 (2011).
17. M. A. Marti-Renom *et al.*, Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct* **29**, 291-325 (2000).
18. T. D. Grant, Ab initio electron density determination directly from solution scattering data. *Nat Methods* **15**, 191-193 (2018).
19. E. F. Pettersen *et al.*, UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612 (2004).