*for **PNAS***
Classification - Major: Biological Sciences
Classification - Minor: Biochemistry

Supplemental Information

Damaged DNA induced UV-damaged DNA-binding protein (UV-DDB) dimerization and its roles in chromatinized DNA repair

Joanne I. Yeh[1,2,*], Arthur S. Levine[3,4], Shoucheng Du[1], Unmesh Chinte[1], Harshad Ghodke[5], Hong Wang[4,5], Haibin Shi[1], Ching L. Hsieh[3,4], James F. Conway[1], Bennett Van Houten[4,5], & Vesna Rapić-Otrin[3,4]

[1]Department of Structural Biology, [2]Department of Bioengineering, [3]Department of Microbiology and Molecular Genetics, [4]University of Pittsburgh Cancer Institute, and [5]Department of Pharmacology and Chemical Biology, University of Pittsburgh School of Medicine, Pittsburgh PA.

Address correspondence to: Joanne Yeh, Department of Structural Biology, University of Pittsburgh School of Medicine, 1036 BST3, 3501 5th Avenue, Pittsburgh, PA 15260, USA, Tel. (412) 648-9027, E-mail jiyeh@pitt.edu

Key Words

UV damage, ubiquitin-proteosome system, chromatin, X-ray crystallography

Materials and Methods

Protein expression and purification: native and SeMet labeled proteins

The protein components of the UV-DDB complex were co-expressed in Sf9 cells and purified through tandem-affinity chromatography with a His-tag on DDB1 and a FLAG-tag on DDB2, following the method that was developed previously (1, 2). Sf9 cells were co-infected with the viruses and incubated for 48 hours at 27 °C in SF900II media (Invitrogen), followed by centrifugation and snap freeze in liquid nitrogen. The frozen pellet was re-suspended in Sf9 lysis buffer 1 (500mM KCl, 50mM potassium phosphate pH 8.0, 10% glycerol, 0.5% NP-40), then rotated for 30 minutes at 4 °C for constant mixing. Centrifugation at 40,000 rpm in a Sorvall T-647.5 rotor for 45 minutes at 4 °C results in clear soluble lysate. The lysate was adjusted to 10mM imidazole and incubated overnight with Ni-NTA superflow resin (Qiagen). The protein bound Ni-resin was washed with 20 column volumes of Sf9 lysis buffer containing 10mM imidazole, then eluted with Sf9 lysis buffer containing 200mM imidazole. Peak fractions were collected and incubated overnight with anti-FLAG M2 affinity gel (Sigma). The protein bound anti-FLAG gel was washed with 20 columns of Sf9 lysis buffer, then eluted with Sf9 lysis buffer 2 (500mM KCl, 50mM potassium phosphate pH 8.0, 10% glycerol) containing 200μg/ml FLAG peptide (Sigma). Peak fractions were collected and concentrated in a 50KDa cut off concentrator (Pall Filtron) to remove the FLAG peptide. The yield of pure UV-DDB complex ranged from 1 to 2 mg/l of Sf9 cells. An EDTA-free protease inhibitor mix (Roche Applied Science) was added to all buffers in the protein purification except in Sf9 lysis buffer 2.

Seleno-L-methionine incorporation

Expression of seleno-L-methionine labeled DDB1 and DDB2 proteins was accomplished by adapting a bacterial expression protocol (3). Sf9 cells were co-infected with baculovirus

encoding each UV-DDB subunit and incubated for 6 hours at 27 °C in SF900II media (Invitrogen). After centrifugation, cells were re-suspended in ESF-921 protein-free methionine-free cell culture medium (Expression Systems LLC) for 8 hours. 250mg/l final concentration of seleno-L-methionine (Acros Organics) was added to each liter of cells and further incubated for 36-40 hours at 27 °C. The same purification procedure was carried out as above except Tris (2-carboxyethyl) phosphine hydrochloride (TCEP) at 1mM final concentration was added to all the buffers. The yield of pure seleno-L-methionine-UVDDB complex ranged from 200-250μg/l of Sf9 cells.

DNA Oligonucleotides: synthesis, purification, and annealing

Single stranded oligodeoxynucleotides were synthesized (Midland Certified Reagent Company Inc.; Midland, Texas) and further purified using anion-exchange chromatography (ProSphere P-WAX; 75x7.5 mm), eluting in a single peak during gradient purification with 25 mM Tris pH 8.5 with 0.02% sodium azide and 0-500mM NaCl. The molecular weights of deoxyoligonucleotides in the purified fractions were confirmed with MALDI-TOF-MS, combined, and concentrated. To form the double-stranded DNA, the purified complementary oligodeoxynucleotides were mixed in a 1:1 molar ratio at room temperature, heated to 90°C, and the samples gradually cooled to room temperature overnight to anneal. After annealing, the oligodeoxynucleotides were further purified using anion-exchange chromatography, as described for the single-strand DNA purification.

After purification, the double-strand DNA samples were buffer exchanged into 20mM Tris HCl, pH 7.5, 2mM $MgCl_2$, 1mM EDTA, 2mM TECP, 5% Glycerol, and 0.02% azide by passing through an Ultracel concentrator three times (Amicon). While several oligodeoxynucleotides containing a tetrahydrofuran (THF) moiety to mimic abasic lesions were synthesized and

purified, the longest oligodeoxynucleotide duplex comprised of 24-base pairs (AP24) containing a central THF lesion site in complex to purified UV-DDB protein produced crystals with well-defined morphologies and better diffraction characteristics. Thus, the structural and functional results described in this study focuses on the UV-DDB-AP24 oligodeoxynucleotide complex. The sequences of the AP24 coding and complementary oligodeoxynucleotide strands are as follows:

AP24 coding strand: 5′-GTCAGCATCG(Abasic)CATCATACAGTCA-3′

Complementary: 5′-TGACTGTATGATGACGATGCTGAC-3′

In addition, for anomalous phasing and to verify positions of DNA strands, brominated AP24 oligodeoxynucleotide were synthesized by replacing all the deoxycytidine by 5-bromodeoxycytidine (^{Br}C) in the 24-mer DNA oligodeoxynucleotide (AP24Br). The brominated single strand oligodeoxynucleotide were purified, annealed, and then purified again, as described for the unhalogenated DNA. The sequences of AP24Br oligodeoxynucleotide are as follows:

AP24Br coding strand: 5′-GT^{Br}CAG^{Br}CAT^{Br}CG(_Abasic) ^{Br}CAT^{Br}CATA^{Br}CAGT^{Br}CA-3′

Complementary: 5′-TGA^{Br}CTGTATGATGA^{Br}CGATG^{Br}CTGA^{Br}C-3′

Electron Microscopy

3µl of sample were pipetted onto a freshly glow-discharge carbon-coated grid, blotted, washed on the surface of a 100µl drop of 1% uranyl acetate stain solution, blotted again and air-dried. Grids were imaged in an FEI T12 microscope operating at 120kV and magnification of 30,000x on film. Micrographs were digitized with a Nikon Super CoolScan 9000 scanner. The ImageJ software (4) was used for image processing of electron micrographs. Image processing to remove background variations, including uneven depth of stain and thickness of the carbon support film, allows the size distribution to be estimated (Figure S1, panels a-f). The dark areas are due to scatter from the uranyl acetate salt, surrounding white stain-excluding areas of

proteins. Preliminary analysis to measure representative areas shows a well-defined peak at ~3500 $Å^2$, corresponding to a spherical particle of ~70Å diameter, consistent with a monomer of the UV-DDB complex. A shoulder is also apparent at twice the area (~7100 $Å^2$), corresponding to a dimer of the UV-DDB, representing ~5% of the particles in the absence of substrate DNA. However, this peak, at ~7100 $Å^2$, is enriched with the addition of damaged DNA (AP24) substrate and approaches 100% when an excess of damaged DNA is present (Fig S1, panels e,f).

AFM sample preparation and imaging

UV-DDB (50 nM concentration of DDB1-DDB2 heterodimer) was incubated with 25 nM undamaged or UV-irradiated 517 bp PCR fragments (25 nM) for 5 mins at 37 °C in a buffer containing 20 mM Tris-HCl (pH 7.5), 150 mM NaCl 0.2 mM $MgCl_2$, and 0.2 mM EDTA. UV-irradiation was done at 20 J/m^2 (254 nm wavelength) on the 517 bp PCR product at 50 μg/ml concentration. The protein-DNA mixtures were diluted 1:5 to 1:10 fold before deposition (25 mM NaOAc, 10 mM $MgOAc_2$, 25 mM Hepes pH 7.5). All samples for AFM imaging were prepared by depositing samples onto a freshly cleaved mica (SPI Supply, West Chester, PA), followed by washing with MilliQ water and drying under a stream of nitrogen gas. All images were collected using a MultiMode V microscope (Bruker, Santa Barbara, CA.) using E scanners in tapping mode. Pointprobe® plus noncontact/tapping mode silicon probes (PPP-NCL, Agilent) with spring constants of ~50 N/m and resonance frequencies of ~190 kHz were used. Images were captured at a scan size of 1 μm × 1 μm, a scan rate of 2-4 Hz, a target amplitude of 0.3 V and a resolution of 512 × 512 pixels.

Volume analysis of AFM images

For AFM volume analysis, dimensions of proteins were measured using Image SXM software (5, 6). AFM volume of a particle was calculated as V= S × (H - B), where V is the

AFM volume, S is the area generated at the base of a protein using "density slice" function of the SXM software (5, 7), H is the average height, and B is the background height. The standard equation relating the AFM volume (V) of a globular protein and its molecular weight (MW) is: V (nm^3) = 0.3856 MW (kDa) - 1.913. This equation is based on AFM volumes of the following proteins and their various oligomeric states: Pot1 (65kDa), PcrA monomer (86.4 kDa), UvrA monomer (105 kDa), *Taq* MutS dimer (181 kDa), UvrA dimer (210 kDa), and *Taq* MutS tetramer (362 kDa).

Quantitative PCR assays

In order to estimate the number of photoproducts induced upon UV exposure, we performed a quantitative PCR assay using untreated or treated template at a concentration of 0.1 ng/ml in a 20 μL reaction. A 517 bp DNA fragment was amplified as described previously (8). Briefly, the cycling conditions are as follows: 75°C for 90 s; 94°C for 5 min; 94°C for 30 s, 57°C for 1 min, and 72°C for 1 min (15 or 16 cycles).

Estimation of the possibility two independent monomeric binding events at vicinal lesions

Calculation of lesion frequency was performed as described before (9). We found that on average each UV irradiated 517 bp fragment of DNA has 1.2 lesions (0.6 photoproducts per strand). Since UV damage is distributed according to a Poisson distribution ($P(k) = \lambda^k e^{-\lambda}/k!$), the percentage of molecules with 0, 1, 2, 3 and 4 or more lesions is approximately 35.37, 35.37, 17.68, 5.89 and 5.69 %, respectively (9).

For the 17.68% of molecules with 2 UV induced lesions, we can calculate the probability of finding 2 lesions in a 30 bp contiguous stretch (corresponding to the footprint of DDB2 on DNA) of DNA as P_2. UV irradiated 517 bp DNA can be modeled as consisting of 517

nucleotides with damaged dinucleotides corresponding to the 6-4PP and CPD lesions. Thus, a DNA fragment with 2 lesions can be treated as a group of 3 objects with the first object being 513 undamaged identical bases and the two lesions corresponding to the other two objects. Therefore, the total number of combinations of 2 lesions is given by *Total outcomes* = $^{515}C_2$ = 132355. The number of combinations of 2 lesions where they are within 30 nucleotides of each other (corresponding to 30 nucleotides which is roughly the footprint of DDB2 dimer on DNA) is calculated as *Favorable outcomes* = $_{\{i=1\}}\Sigma\ ^{1}C_1{}^{(515-i)}C_1$ number of combinations of the two lesions so that the number of nucleotides between them is 0, 1, 2 and so on until 27. Thus for a DNA fragment containing two lesions, the probability of both of those lesions lying within the footprint of DDB2 dimer is P_2 = favorable outcomes/total outcomes = 0.10. This corresponds to a total of about 1.77% of all molecules with 2 lesions within the foot print of the DDB complex.

Similarly, a 517 bp fragment of DNA containing 3 lesions can be treated as 4 objects – 511 non damaged bases, and 3 lesions. We can calculate the probability of finding at least 2 lesions within the footprint of DDB for DNA molecules containing 3 lesions as $P_3 = 1-\underline{P}_3$, where $\underline{P}_3$ is the probability of the 3 lesions always being outside of the footprint of DDB2 dimer. $\underline{P}_3$ = favorable outcomes where the lesions do not lie within 30 bases of each other/total outcomes. The total number of combinations of 3 lesions is *Total outcomes* = $^{513}C_3$ = 22369536. The number of favorable outcomes = $^{\{513-28\}}{}_{\{i=1\}}\Sigma\ ^{i}C_1{}^{(513-28-i)}C_1$.Thus $\underline{P}_3$ = 0.85, and P_3 = 0.15. Thus, about 0.88% of all molecules have 3 lesions within the foot print of the DDB complex.

We can similarly calculate the probability of finding at least 2 lesions within the footprint of DDB for DNA molecules containing 4 lesions or more however, this is a sufficiently small fraction of molecules amounting to just 1.5% for molecules with 4 lesions (with the percentage of molecules with more than 4 lesions being even smaller) of molecules that we can neglect it

without significantly affecting our analysis. Based on these numbers, we expect approximately 2.5% of all molecules to contain multiple lesions within the footprint of the DDB complex.

UV-DDB/DNA crystallization

For crystallization, the UV-DDB complex was mixed with the purified AP24 duplex DNA in a molar ratio of 1:3 (UV-DDB:DNA) in 20mM Tris pH 7.5, 2mM $MgCl_2$, 1mM EDTA, 2mM TECP, 5% Glycerol, and 0.02% azide. The sample was concentrated to about 2.5 mg/ml (UV-DDB) using an Ultracel concentrator. Numerous crystallization screening trays were set up at 4°C and the most promising conditions were optimized using a hanging drop diffusion method with a volume ratio of 1:1 (µl) of protein solution to reservoir. For data collection, crystals were transferred into a solution containing a cryoprotectant, typically comprised of the crystallization solution augmented with 22% ethylene glycol, then flash cooled in liquid nitrogen.

Preliminary small crystals were grown from a stock containing native human UV-DDB protein that was incubated with small excess of AP24 DNA oligomer. To ascertain that AP24 was stoichiometrically bound to UV-DDB in the crystallization condition, small crystals were harvested, washed, dissolved, and confirmed by gel-electrophoresis. The early crystals exhibited multiple morphologies and diffracted weakly to 8Å. Further optimization using additive protocols (10) resulted in crystals with better defined crystal habit and morphology and were subsequently used for streak- and micro-seeding. After iterative cycles of optimization, single prismatic crystals exhibiting gradually improved X-ray diffraction characteristics were obtained. These crystals were used as seeds for further optimization using the matrix seeding method (11-13), identifying several new promising conditions. The final crystallization condition yielding single crystals used for data collection contained 0.25M potassium thiocyanate, 0.08M Bis-Tris propane pH 6.5, 28% w/v PEG 3350, and 2% 1,5-pentanediol. The UV-DDB complex

crystallized in a monoclinic lattice, with a screw axis along the unique two-fold rotation axis (space group $P2_1$) with unit cell dimensions of a=76.736 Å, b=70.877 Å, c=191.448 Å, α=90.00°, β=99.68°, γ =90.00°. APS synchrotron diffraction data to 2.85 Å resolution were collected on the 24ID beamline.

For phasing, Se-methionine substituted UV-DDB (DDB1-SeMet and DDB2-SeMet) were co-crystallized with AP24 (1:3 of UV-DDB to DNA) from a solution containing 0.35M di-ammonium tartrate and 30 %(w/v) PEG 3350, and nucleated using seeds transferred by the streak seeding method. The SeMet-substituted UV-DDB/AP24 complex crystallized in an orthorhombic lattice (spacegroup $P2_122_1$), with similar unit cell parameters for two of the axis while the third unique axis doubled in length. The *ab-inito* selenomethionine SAD data was phased to 3.2 Å (Table 1, Table S1). with a solvent content of ~56% with one molecule in the asymmetric unit. The DDB1-DDB2 sequence has a total of 37 methionine residues. For anomalous phasing, SAD datasets were collected on selenomethionine substituted crystals and the datasets were merged for high redundancy. SeMet SAD data collections were done at SER-CAT 22ID and 22BM beamlines at APS. The SAD datasets were collected at peak energy wavelength, optimizing redundancy (Table S1). For all datasets, the crystals were translated throughout the diffraction experiments, after collection of a small wedge of data, to minimize effects of radiation damage.

Brominated DNA (AP24Br) was also synthesized for phasing, substituting cytosine with 5-bromo-dC, purified, annealed, and purified as described earlier. Purified UV-DDB was co-crystallized with AP24Br. Similar to the crystallization of UV-DDB-AP24, UV-DDB-AP24Br initial crystallization hits also only grew tiny crystals with poor morphology. Further additive screening with a condition containing 0.2M sodium fluoride, 0.1M Bis Tris propane pH 6.5, and

33% w/v PEG 3350 produced single diffracting crystals. The best crystals were obtained with the additive, 0.08M GSH (L-Glutathione reduced) and GSSG (L-Glutathione oxidized). Data sets were collected at the Argonne Photon Source (Chicago, Illinois), on the SER-CAT and GM/CA beamlines and at the Stanford Synchrotron Radiation Lightsource (Palo Alto, CA), on beamlines BL7-1 and BL9-1. Due to the limited resolutions of diffraction of the AP24Br datasets (nominally 3.5-3.6 Å), the Br-datasets were used only in the early phase of model building, primarily to verify the placement and orientations of the AP24 DNA strands in the early models.

X-ray data collection, structure determination, and model building

Two datasets collected from crystals grown using SeMet-substituted UV-DDB protein were merged during processing, improving overall completeness and redundancy (Table S1). All datasets were processed using the HKL2000 suite (14). For the anomalous data, the intensity measurements for the Friedel pairs were separated at the scaling stage in HKL2000. To resolve the enantiomorphic ambiguity (i.e., hand of the substructure atom configuration) of the SAD phases, the phases were combined with those calculated from a model. Partial model phasing using the monoclinic P21 UV-DDB-AP24 structure was initially done in Phaser (15). Phase combinations of the calculated phases together with the SeMet SAD data were done in Phenix (16, 17), producing well-defined electron density maps. In one UV-DDB subunit, a total of 27 selenium sites were found, with 21 sites having occupancy greater than 0.6. The initial FOM from the combined phasing was 0.76 as compared to 0.27 using the SAD data alone in Solve (18). Attempts to phase directly by molecular replacement using the human DDB1 and zebrafish DDB2 (pdb accession numbers 3EI2-3EI4) resulted in poor rotation and translation function values, with FOM of ~0.25 and poorly defined electron density maps with large regions of discontinuous densities and the AP24 oligodeoxynucleotide. The maps calculated using

combined phasing methods showed clear electron density for the AP24 DNA molecule, located close to the DDB2 domain (Figure S7). These maps were used for model rebuilding in Coot (19). Once the relative orientations of the DDB1, DDB2, and AP24 DNA were determined, these were cross-validated using the SeMet anomalous dataset. This orthorhombic crystal form (pdb ID code 4E5Z) independently confirmed the NT-domain fold and cross-validated the structure determined in the monoclinic lattice (pdb ID code 4E54) (refinement statistics for both crystal structures are shown in Table 1).

Extensive regions of the UV-DDB were traced and built *de-novo*, including loop regions connecting the beta sheets in DDB1, the β-wings, and terminal regions of both DDB1 and DDB2. The NT-domain of DDB2, comprised of residues 20-100, was gradually built into experimental electron density maps as the earlier published structures were missing this NT region. The NT-region of DDB2 (residues 20-100) was iteratively extended, locally then globally refined, validated by iterative composite omit maps using the 2.85Å dataset. Iterative refinement cycles included residue-by-residue fitting followed by energy minimization and grouped/isotropic B-factor refinement over the entire complex. In addition to the NT-domain of DDB2, several regions were verified for model accuracy by generating simulated annealing omit maps in Phenix; these included the newly extended NT-helical α-paddle and β-wing regions, several loops at the DDB1-DDB2 interface, and the BPB domain of DDB1.

For AP24 DNA, density was seen for all 24 bases for both strands. Also, for the damaged DNA strand no extra density was seen in the difference maps around the tetrahydrofuran (THF) legion, signifying the absence of a nucleobase. During DNA fitting, the planarities of complimentary bases were restrained initially, with loosening of restraints as refinement progressed except for a 2 base-pair window around the abasic site, where the planarity is

disrupted. As seen in earlier structures of UV-DDB-DNA complexes, DDB2-induced DNA kinking is also seen in our UV-DDB-AP24 complex structures. The UV-DDB-AP24 complex has been refined to Rwork/Rfree values of 0.22/0.24 (monoclinic) and 0.25/0.26 (orthorhombic), with refinement statistics shown in Table 1 & Table S1.

Dynamic Light Scattering (DLS) Analysis

The DLS analysis was performed on a DynaPro (Wyatt Technology) molecular-sizing instrument equipped with a Plate Reader (Protein Solutions). A 20 μL UV-DDB sample was passed through a 0.20 μm filtering assembly into the sample chamber of the DynaPro. Data collection and analysis utilized Dynamics 6.0 software package, as originally described (20). For analyses, samples of UV-DDB were prepared using buffer (10 mM HEPES pH 7.4 and 0.15 M NaCl) to adjust total protein concentrations immediately before light scattering measurements. Data were collected under identical experimental conditions before and after the addition of damaged DNA substrate (AP24). The same molar ratio of 3:1 AP24:UV-DDB as that used for crystallization was maintained. The particle sizes and molecular weights (MW) of UV-DDB in the presence and absence of damaged DNA (AP24), calculated from the DLS data collected at six different protein concentrations, are summarized in Table S3.

Surface Plasmon Resonance (SPR) analysis

All SPR experiments were performed by using a BIAcore 3000 biosensor and Sensor Chip CM5 (GE Healthcare) at 6 ºC including maintaining the sample holders at this temperature with a circulating water-bath. All of the AP24 samples were diluted into running buffer (10 mM HEPES, pH 7.5, 0.15 M NaCl), with sample concentrations from 1.4 nM to 1.0 μM. All measurements were performed in series, with one channel dedicated as a control with its surface generated as similar as possible to the sample surface channel (i.e., channel

immobilized with UV-DDB). Immediately following injection of AP24 (i.e., damaged DNA substrate) through the UV-DDB immobilized sample channel, AP24 was injected in the reference channel and sample channel, at flow rates of 30 μL/min for 2.5 min. Background signals caused by refractive index changes and non-specific surface interactions were recorded in tandem. The sensorgram signals from the reference channel were subtracted from the sample channel (i.e., UV- DDB immobilized channel) to obtain the overall signal corresponding to the binding of the AP24 to UV-DDB. At the end of the association phase, the flow rate in the absence of the AP24 substrate was maintained to monitor dissociation for duration of 5 min. A 2M NaCl buffered solution regenerated the surface.

Data analyses were performed using the BIAevaluation software to model the binding of AP24 to UV-DDB in the monomeric and dimeric forms. Preliminary assessment evaluating the feasibility of the SPR approach to detecting changes in protein-substrate (AP24) binding parameters were conducted under steady-state conditions. Additional optimization of SPR experimental conditions and immobilization protocols led to two separate methods for the immobilization of the dimeric form of UV-DDB (further elaborated below). All measurements were replicated in triplicate and cross-validated when possible for comparison and cross-validation of several data sets. SPR datasets collected for the binding of AP24 to the monomeric form of UV-DDB were modeled by fitting the data to a Langmuir adsorption model. Preliminary SPR data at high UV-DDB concentrations suggested that the binding of AP24 to the dimeric form of UV-DDB exhibited interactions that are more complex. Two different protocols for the immobilization of the dimeric UV-DDB were used for the SPR analysis, generating two sets of data, each replicated in triplicate, for cross-validation. To verify that predominantly monomeric or dimeric forms of UV-DDB were immobilized at specific protein concentrations, DLS analyses were done over a range of concentrations as described in the preceding section, in the absence

and presence of AP24. Accordingly, at 20 μg/mL predominantly monomeric or at 800 μg/mL dimeric UV-DDB forms were immobilized in separate channels on research-grade carboxymethyl (CM5) chip surfaces via standard EDC/NHS-mediated amine coupling procedures (21). Briefly, the carboxymethyl dextran surface was activated using freshly a prepared aqueous solution containing 0.2 M EDC and 0.05 M NHS. Once the carboxymethyl (CM5) chip surface was activated, injection of purified UV-DDB covalently linkage formed between accessible nucleophilic groups (primary and secondary amines) of UV-DDB to the freshly formed succinamide groups on the chip surface. The same protocol was repeated by injecting purified and concentrated UV-DDB at 800 μg/mL, immobilizing the dimeric form in a separate channel. Unbound protein was removed and unreacted surface sites on the CM5 chip were capped by injecting 1M ethanolamine-HCl (pH 8.5). A series of AP24 association and dissociation experiments to monitor the interactions of AP24 to monomeric and dimeric UV-DDB. Contributions from non-specific interactions to the measured SPR signals were subtracted dynamically. Consistent experimental parameters and conditions were maintained between the reference and sample channels.

An alternative method to generate the dimeric form in-situ on the SPR surface was used for determining parameters for the binding of AP24 to monomeric and dimeric forms of UV-DDB in series. Using this approach, data from the association phase for the binding of AP24 to the monomeric form of UV-DDB was measured, and complete dissociation of AP24 using 2M NaCl to regenerate unbound (apo) monomers of UV-DDB. To form the dimeric form of UV-DDB in-situ, highly concentrated UV-DDB was injected at 3mM, a concentration several times greater than the estimated affinity of dimerization, generating dimeric UV-DDB through innate oligomerization tendencies. The dimerization was allowed to proceed for 150 s and the stability

of the subsequent dimeric UV-DDB formed in-situ on the SPR chip surface was assessed under low flow rates while monitoring the sensorgram for dissociation. The submicromolar dissociation constant (KD) estimated for the monomer-dimer UV-DDB association is supported by the stable baseline that indicated that the dimeric form of UV-DDB was stably maintained. The binding of AP24 to dimeric UV-DDB was measured by injecting UV-DDB-ligand at 3.0, 1.0, 0.33, 0.11, 0.037, and 0.012 mM pre-incubated with AP24 (at the ratio of 1:3) through the UV-DDB affixed channel, at a flow rate of 30 µl/min. The association was allowed to proceed for 150 s, and dissociations of the complex monitored for 300 s and the data used to calculate the rates and affinities of AP24 binding (Table S3).

The values obtained for the dimeric form of UV-DDB generated in-situ to those calculated from the sets of data collected from immobilizing the dimeric UV-DDB at a high concentration (800 µg/mL) are comparable. Accounting for the heterologous population in the rates and affinity calculations gave generally comparable values and trends. The contributions from the mixed monomer-dimer population were accounted for by using the state distributions obtained from DLS analysis**.** Using BIAevaluation software (version 4.1) modeling according to heterogeneous ligand-parallel reaction to fit the data resulted in low residuals, indicating statistically valid fit. The kinetics and affinity constants estimated for AP24 in the monomeric and dimeric states of UV-DDB are tabulated in Table S3. It should be noted that additional experimental analysis are required to fully validate these preliminary results.

Expanded Discussion

Dimer Interface Adjacent to Damaged DNA Binding Site in DDB2

The primary DNA binding surface is located at the narrow end of the β-propeller, opposite the DDB1 interaction region of DDB2 (Figure 2). The bound DNA molecule spans the

surface of DDB2, offset from the center of the 7-bladed β-propeller (Figure S2A). This offset can now be explained in terms of the spatial constraints arising from the dimerization interface. To accommodate the steric requirements allowing for dimerization of DDB2 along with binding of damaged DNA substrate (Figure S2B), the neighbouring intermolecular contact surfaces are located diametrically across a molecular face of the β-propeller domain of DDB2, leading to the offset (Figure S2C). The location of the dimer interface on the same molecular surface as the DNA binding site allows for cooperativity and coordination of oligonucleotide binding and dimerization.

Using the nomenclature introduced previously (22), the damage site is denoted as "D_{+1}" with the nucleotides 3' of the damaged site denoted as D_{+n} and 5' as D_{-n}, where the number refers to the location of the nucleotide relative to the damage site at D_{+1}. Overall, localized deformation at the abasic site forms a gap of ~15 Å and unwinds the DNA by ~18 degrees, radiating from the lesion (Figure S3). This perturbs the stacking of the bases immediately opposite the abasic site and the adjacent upstream nucleotide but limited in range to within two-nucleotides of the lesion.

DNA damage is detected by insertion of the β-loop extending from strand-5 of the β-propeller domain of DDB2 into a gap in the duplex formed by CPD, 6-4PP, or abasic lesions, which all produce similar nucleotide deformations in the duplex (22). This insertion loop is comprised of three highly conserved residues -- Phe334, Gln335, and His336 – with each residue forming specific contacts to the bases opposite the lesion (Figure S3C). The minor groove at the lesion site is occupied by the abasic ribose moiety, positioned through the extra-helical flipping of the adjacent nucleobase (dC12), which appears to drive the distortions at the lesion as no significant interactions to DDB2 are found. The β-loop of strand 3 of DDB2 defines the perimeter of the site occupied by the nucleobase (dC12) immediately upstream of the lesion.

This base is an extra-helical configuration, which enlarges the gap initially produced by the missing nucleobase (THF11). The resulting gap is filled by the imidazolium ring of DDB2 His336 on the 5' side and by Gln335 on the 3' side, limiting the lesion-induced distortion from being propagated. The extrahelical conformation of dC12 is maintained by a combination of H-bonds, hydrophobic, and π-stacking interactions between the cytosine base and DDB2 residues, particularly Ile200, Asn201, and Trp203 (Figure S3). The plane of the aromatic ring of Trp203 forms the shallow base-binding plateau with Met177 and Ile200 bordering the site, which can accommodate both pyrimidine and purine nucleobases. Additional sequence independent stabilization at the flipped nucleotide next to the lesion is afforded through electrostatic contacts between the phosphoribose-backbone of the DNA to Arg and Lys residues of DDB2.

Immediately 5' of the flipped dC12, dA13 remains in the duplex, stacking over the carboxamide moiety of Gln335, which replaces the nucleobase of dC12. Gln335 also forms H-bonds to His336, stacking the imidazole ring over the purine ring of dG10, so that the canonical base-stacking interactions 5' at the D_{+2} and 3' at the D_{-1} regions are maintained in the damaged strand. The positively charged guanidinium groups of Arg112 bridge the two strands at D_{-1}, D_{-2}, maintaining inter-strand base-pairing contacts on the 5' side of the lesion (Figure S3C).

Dimeric UV-DDB Binds Damaged DNA with Higher Affinity than in the Monomeric State

The DLS experiments clearly show that the binding of damaged DNA substrate (AP24) induces UV-DDB dimerization, even at the lowest protein concentration analyzed, 100 μg/mL concentration (which is 25X more dilute than the stock concentration used for crystallization). As the plots show (Figure S8), under the solution conditions for the DLS analysis (identical to SPR experimental conditions), in the absence of damaged DNA, a maximum of ~15% of UV-DDB was dimeric at the highest protein concentration studied, 1.25 mg/mL. In contrast, in the

presence of AP24, only dimeric UV-DDB was found at 1.25 mg/mL and less than 5% were monomeric even at the most dilute protein concentration analyzed, 0.1 mg/mL. Taken together with the kinetics and binding affinities determined by SPR, both performed under identical solution conditions, the binding of damaged DNA induces dimerization of UV-DDB, which further enhances the tightness of binding, ~doubling the rates of association and reducing the dissociation rates by 50%. These results are consistent with the findings from X-ray crystallography, EM, and AFM, supporting the role that dimerization plays in modulating binding affinity, helping to assure fidelity of damage detection. The actual degree to which kinetic and affinity parameters are altered by dimerization need to be examined using UV-damaged DNA substrates – these analyses reflect the consistent trend found between solution and structural findings involving AP deoxyoligonucleotides.

Functional implications of the dimeric state of UV-DDB

Our analysis of dimeric UV-DDB provides conformational insights on potential associations between a substrate receptor and substrate, suggestive of dimerization functioning in regulating overall activity of the E3 ligase complex. Based on the additional extensive contacts formed by the dimeric UV-DDB, this new conformation can modulate substrate affinities on multiple levels, serving to allosterically regulate the substrate-receptor complex. Similar to the transcriptional factor, (Ets-1)2–S-EBS complex (Ets-1) (23), the dimeric state of UV-DDB is stabilized by DNA binding, inducing conformational ordering in the NT domain and further generating additional DNA-binding surfaces. In the Ets-1, dimerization is key to relieving auto-inhibition, whereby ternary complex formation of Ets-1 with itself initiates the conformational transition to a high-DNA-binding affinity state (23). In the UV-DDB, damaged DNA binding triggers the NT domain of DDB2 to adopt an α-helical paddle motif, presenting new DNA

binding sites. Additional DNA-contact sites are generated once the β-wing conformation is induced, concomitant with dimerization (Figure 2).

The β-wing loop of DDB2 is located at the interface of two DNA molecules in the dimeric UV-DDB, so that multiple unique sites of DNA contacts are found in the dimeric UV-DDB, interactions that are absent in the monomeric state. These interactions are centered at the NT-α-helical paddle and β-wing regions of DDB2 and when analyzed together resemble a "winged helix" motif that has been found in numerous DNA-binding proteins (24).

Winged helix proteins share a related DNA-binding motif, combining beta loops ('wings'), alpha helices and beta-sheets to modulate DNA contacts. In these proteins, the helices form various levels of DNA contacts, from sequence-specific major groove insertions to base-independent electrostatically-mediated contacts to the deoxyribose-phosphate backbone of the DNA. Analogous to those found in other winged-helix DNA binding proteins, the β-wings in UV-DDB form direct contacts to the backbone atoms of the DNA (Figure 2). Additionally, the conformation and the apparent function of the β-wings of DDB2 in dimeric UV-DDB resemble those shown in the transcription factors, Ets-1, and the tripartite factor X, RFX, by linking and modulating nucleotide binding affinities with dimerization (23-25). In these winged-helix proteins, patches of hydrophobic residues are exposed causing conformational changes to present new protein-protein interaction surfaces and inducing dimerization as a function of nucleotide binding.

The winged-helix DNA-binding motif in the dimeric UV-DDB is obtained by combining the α-paddle with the β-wing motifs in the dimeric DDB2 (23-26). The orientation of the NT-α-paddle domain of DDB2 relative to the DNA is similar to those formed between the helical

segments of Ets-1 to the DNA (PDB accession code 2nny). Both proteins contain clusters of acidic and basic amino acids that form charge and dipolar interactions on the helical surfaces facing the DNA backbone.

Using our structural information to align the NT domain of DDB2 to the H-T-H winged-type DNA-binding motif in Ets-1 and RFX results in a high conservation of sequence homology (>70% identity) and in the 3D-fold. When only the primary sequence of monomeric DDB2 was used in various prediction programs to identify protein function, including ProFunc (26), only the WD40 domain was identified whereas the programs were unsuccessful in recognizing other functional motifs in DDB2. However, defining specific domain boundaries using the dimeric UV-DDB structure and combining the sequences of these as composite sites, allowed the successful identification of DNA binding sites in DDB2. Thus, these structural analyses support the findings of the crystal structure, EM, and AFM studies, verifying that non-contiguous regions of DDB2 can adopt DNA-binding motifs when assembled in the dimeric state. Consequently, in DDB2, the NT-helical and β-wing domains contain conserved DNA binding sequences that allow DDB2 to adopt high-affinity DNA binding motifs. Segregating these topologies so that the high-affinity winged-helix motif is formed only upon dimerization modulates DNA affinities, enabling apparently incompatible kinetic and thermodynamic oligonucleotide-binding characteristics to be encoded within a single protein.

Supplementary Tables

Table S1: Data Statistics for SeMet UV-DDB

	UV-DDB-AP24 SeMet, Dimer Dataset 1
Data collection	
X-ray source	23ID-B, APS
Bravais Lattice	Orthorhombic
Cell Dimensions	
a, *b*, *c* (Å)	72.42, 76.50, 389.74
α, β, γ (°)	90, 90, 90
Resolution (Å)	3.22
# Unique reflections	30981
R_{merge} (3.2-3.31 Å)*	0.117 (0.358)
$I / \sigma I$ (3.2-3.31 Å)	10.9 (3.1)
Completeness (%) (3.2-3.31 Å)	93.8 (71.3)
Redundancy (3.2-3.31 Å)	5.6 (3.6)

*Values in parenthesis are for the highest resolution shell (3.2-3.31 Å)

Table S2. Dynamic Light Scattering: oligomeric states of UV-DDB in the presence and absence of damaged DNA (AP24)

Sample	Molecular Weight of UV-DDB (kDa)	% Monomer 200 kDa	% Dimer 400 kDa	Oligomeric State
No Damaged DNA (AP24) **UV-DDB [0.1 mg/mL]**	201	>98.5	<1.5	Monomer
		Weighted <MW> kDa 205		
No Damaged DNA (AP24) **UV-DDB [0.4 mg/mL]**	209	~95	~5	Monomer-Dimer
		Weighted <MW> kDa 221.3		
No Damaged DNA (AP24) **UV-DDB [1.25 mg/mL]**	261	~85	~15	Monomer-Dimer
		Weighted <MW> kDa 245.0		
+ Damaged DNA (1:3 AP24*) **UV-DDB-AP24 [0.1 mg/mL]**	422	<1	>99	Dimer
		Weighted <MW> kDa 397.5		
+ Damaged DNA (1:3 AP24*) **UV-DDB-AP24 [1.25 mg/mL]**	504	0	100	Dimer
		Weighted <MW> kDa 400.0		

Table S3. Surface Plasmon Resonance: relative rates and affinities of damaged DNA (AP24) binding for monomeric and dimeric UV-DDB

Oligomeric State of UV-DDB	Parameters					
	k_a (1/Ms)	k_d (1/s)	K_A (1/M)	K_D (M)	Chi^2	ΔK_D (Fold Increase ↑) Dimer vs Mono
SPR Data Set 1						
AP24 to UV-DDB monomer	2.38×10^{5}	5.97×10^{-3}	3.98×10^{7}	2.5×10^{-8}	0.82	**4X (3.52)**
AP24 to UV-DDB dimer	4.52×10^{5}	3.21×10^{-3}	1.41×10^{8}	7.10×10^{-9}	1.01	
SPR Data Set 2						
AP24 to UV-DDB monomer	2.16×10^{5}	6.24×10^{-3}	3.46×10^{7}	2.88×10^{-8}	1.19	**4X (3.98)**
AP24 to UV-DDB dimer	4.33×10^{5}	3.13×10^{-3}	1.38×10^{8}	7.22×10^{-9}	1.10	
Summary Monomer Data Sets 1&2						
AP24 to UV-DDB monomer	2.38×10^{5}	5.97×10^{-3}	3.98×10^{7}	2.5×10^{-8}	0.82	
AP24 to UV-DDB monomer	2.16×10^{5}	6.24×10^{-3}	3.46×10^{7}	2.88×10^{-8}	1.19	
Monomer <Average>	*2.27×10^{5}*	*6.11×10^{-3}*	*3.72×10^{7}*	*2.69×10^{-8}*	--	
Summary Dimer Data Sets 1&2						
AP24 to UV-DDB dimer	4.52×10^{5}	3.21×10^{-3}	1.41×10^{8}	7.10×10^{-9}	1.01	
AP24 to UV-DDB dimer	4.33×10^{5}	3.13×10^{-3}	1.38×10^{8}	7.22×10^{-9}	1.10	
Dimer <Average>	*4.43×10^{5}*	*3.17×10^{-3}*	*1.40×10^{8}*	*7.16×10^{-9}*	-	
Overall Δ in Kinetics & Affinities of Damaged DNA (AP24) Binding Parameters in Dimeric Relative to Monomeric States of UV-DDB	***2X***	***½-1X***	***2-4X***	***¼-½X***	*--*	***Association rate 2X faster Dissociation rate 2X slower Overall ~4X Higher Affinity in dimeric state (relative to monomeric)***

** Molar ratio of UV-DDB to AP24 is 1:3, maintaining the stoichiometric ratios used for crystallization and EM analysis*

Table S4. Summary of UV-DDB molecular parameters

[UV-DDB] & Method	Radius (nm)	Area (nm^2)	Volume (nm^3)	Molecular Weight (kDa)	Oligomeric State(s)
No DNA or Bound to Undamaged DNA					
DLS: No DNA					
UV-DDB [UV-DDB]=0.5 µM	3.3			188	Monomer
UV-DDB [UV-DDB]=2.2 µM	3.7			202	Monomer
EM: No DNA					
UV-DDB [UV-DDB]=0.5 µM	3.5	35			Monomer
AFM: +Nondamaged DNA					
UV-DDB 25 nM undamaged DNA; [UV-DDB]=50 nM			~81 ± 10	184 ± 23	Monomer
Bound to Damaged DNA					
DLS: +AP24 Abasic DNA					
UV-DDB-AP24 1:3 AP24; [UV-DDB]=0.5 µM	6.7			403	Dimer
UV-DDB-AP24 1:3 AP24; [UV-DDB]=2.2 µM)	7.0			433	Dimer
EM: +AP24 Abasic DNA					
UV-DDB-AP24 1:3 AP24; [UV-DDB]=0.5 µM	~5	71			Dimer
AFM: +UV-irradiated 517 bp PCR DNA					
UV-DDB-AP24 1:2.5 UV-DNA; [UV-DDB]=50 nM			~133	349.9	Dimer+ 1 Duplex DNA
			~139	365.4	Dimer+ 2 Duplex DNA

Table S5: XPE Mutations

XP-E Patient	Mutation	Biochemical DDB2-DDB1	Biochemical DDB2-UV-DNA	References
XP2RO	R273H	affected	affected	(2, 27, 28)
XP25PV	D307Y	affected	affected	(2)
GM01389	L350P delN349	affected	affected	(2) (29)
XP82TO	K244E		affected	(2, 27, 28)

Supplemental References

1. Wittschieben BO, Iwai S, & Wood RD (2005) DDB1-DDB2 (xeroderma pigmentosum group E) protein complex recognizes a cyclobutane pyrimidine dimer, mismatches, apurinic/apyrimidinic sites, and compound lesions in DNA. *J Biol Chem* 280(48):39982-39989.
2. Rapic-Otrin V*, et al.* (2003) True XP group E patients have a defective UV-damaged DNA binding protein complex and mutations in DDB2 which reveal the functional domains of its p48 product. *Hum Mol Genet* 12(13):1507-1522.
3. Cronin CN, Lim KB, & Rogers J (2007) Production of selenomethionyl-derivatized proteins in baculovirus-infected insect cells. *Protein Sci* 16:9.
4. Abramoff MD, Magelhaes PJ, & Ram SJ (2004) Image Processing with ImageJ. *Biophotonics International* 11(7):36-42.
5. Ratcliff GC & Erie DA (2001) A novel single-molecule study to determine protein--protein association constants. *J Am Chem Soc* 123(24):5632-5635.
6. Wang H, Yang Y, & Erie DA (2007) Characterization of protein-protein interactions using atomic force microscopy *Protein Interactions Biophysical approaches for the study of complex reversible systems,* ProteinReviews, ed Schuck P (Springer Science+Business Media, LLC), Vol 5, pp 39-78.
7. Wang QE*, et al.* (2007) Ubiquitylation-independent degradation of Xeroderma pigmentosum group C protein is required for efficient nucleotide excision repair. *Nucleic Acids Res* 35(16):5338-5350.
8. Wang H, Tessmer I, Croteau DL, Erie DA, & Van Houten B (2008) Functional characterization and atomic force microscopy of a DNA repair protein conjugated to a quantum dot. *Nano Lett* 8(6):1631-1637.
9. Ayala-Torres S, Chen Y, Svoboda T, Rosenblatt J, & Van Houten B (2000) Analysis of gene-specific DNA damage and repair using quantitative polymerase chain reaction. *Methods* 22(2):135-147.
10. Yeh JI, Beale SI (2007) Calorimetric Approaches to Characterizing Effects of Additives on Protein Crystallization. *Crystal Growth and Design* 7(11):2134-2139.
11. D'Arcy A, Villard F, & Marsh M (2007) An automated microseed matrix-screening method for protein crystallization. *Acta Crystallogr D Biol Crystallogr* 63(Pt 4):550-554.
12. Ireton GC & Stoddard BL (2004) Microseed matrix screening to improve crystals of yeast cytosine deaminase. *Acta Crystallogr D Biol Crystallogr* 60(Pt 3):601-605.
13. Walter TS*, et al.* (2008) Semi-automated microseeding of nanolitre crystallization experiments. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 64(Pt 1):14-18.

14. Otwinowski Z, Minor, W. (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods in Enzymology: Macromolecular Crystallography, part A* 276:307-326.
15. McCoy AJ, *et al.* (2007) Phaser crystallographic software. *J Appl Crystallogr* 40(Pt 4):658-674.
16. Terwilliger TC, *et al.* (2009) Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallogr D Biol Crystallogr* 65(Pt 6):582-601.
17. Terwilliger TC, *et al.* (2008) Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr D Biol Crystallogr* 64(Pt 1):61-69.
18. Terwilliger TC & Berendzen J (1999) Automated MAD and MIR structure solution. *Acta Crystallogr D Biol Crystallogr* 55(Pt 4):849-861.
19. Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1):2126-2132.
20. Moradian-Oldak J, Leung W, & Fincham AG (1998) Temperature and pH-dependent supramolecular self-assembly of amelogenin molecules: a dynamic light-scattering analysis. *J Struct Biol* 122(3):320-327.
21. Johnsson B, Lofas S, & Lindquist G (1991) Immobilization of proteins to a carboxymethyldextran-modified gold surface for biospecific interaction analysis in surface plasmon resonance sensors. *Anal Biochem* 198(2):268-277.
22. Scrima A, *et al.* (2008) Structural basis of UV DNA-damage recognition by the DDB1-DDB2 complex. *Cell* 135(7):1213-1223.
23. Lamber EP, *et al.* (2008) Regulation of the transcription factor Ets-1 by DNA-mediated homo-dimerization. *EMBO* 27:2006-2017.
24. Rhodes D & Burley SK (2000) Protein–nucleic acid interactions. *Current Opinion in Structural Biology* 10:75–77.
25. Nekrep N, *et al.* (2002) Mutation in a winged-helix DNA-binding motif causes atypical bare lymphocyte syndrome. *Nat Immunol* 3(11):1075-1081.
26. Laskowski RA, Watson JD, & Thornton JM (2005) ProFunc: a server for predicting protein function from 3D structure. *Nucleic Acids Res* 33:W89-W93.
27. Nichols AF, Ong P, & Linn S (1996) Mutations specific to the xeroderma pigmentosum group E Ddb- phenotype. *J Biol Chem* 271(40):24317-24320.
28. Shiyanov P, *et al.* (1999) The naturally occurring mutants of DDB are impaired in stimulating nuclear import of the p125 subunit and E2F1-activated transcription. *Mol Cell Biol* 19(7):4935-4943.

29. Nichols AF, *et al.* (2000) Human damage-specific DNA-binding protein p48. Characterization of XPE mutations and regulation following UV irradiation. *J Biol Chem* 275(28):21422-21428.

Supplementary Figure Legends

Figure S1. Evidence of dimerization of UV-DDB from particle size analysis from negative stain electron micrographs.

A representative area of the UV-DDB sample without DNA is shown (a) raw; (b) band-pass filtered to preserve features from 5-18 nm in dimension; (c) after application of a local "rolling ball" filter (radius 9 nm) to further remove variations in the background gradient; and (d) thresholded to yield countable areas (white). This procedure was applied to all micrographs and results were combined into a total histogram for each sample. (e) A histogram of particle areas estimated from a single negative stain micrograph shows a peak at ~36 nm^2, corresponding to a circle of diameter ~7 nm that is consistent with a monomer of the UV-DDB1-DDB2 complex, and a shoulder at ~72 nm^2 that represents a minor population of dimers. Examples of monomer-sized areas are indicated with arrowheads, and dimers with double-arrowheads, in (d). The ImageJ software (4) was used for image processing. (f) A gallery of individual particles from different techniques of electron microscopy, as indicated. Some negative stain images are as large as monomers of UV-DDB ("1") while others are twice ("2") as large.

Figure S2: Electrostatic potential surfaces of dimeric UV-DDB2 in a complex with AP24 DNA

Electrostatic potential surfaces generated in PyMol, with positive potentials displayed in blue, negative in Electrostatic potential surfaces generated in PyMol, with positive potentials displayed in blue, negative in red, and uncharged in white. (A) The BPA domain of DDB1 displays a large hydrophobic region (white surfaces, outlined in a dotted yellow line) on the surfaces facing the β−propeller domain of DDB2. In contrast, the BPC domain of DDB1 presents strong negative

charge clusters (surfaces in red, outlined in a purple dotted line) facing the NT α−paddle domain of DDB2. (B) DDB2 displays strong positive charge clusters (in blue) facing DDB1's BPC domain (outlined in a green dotted line) and at the DNA binding sites. These electrostatically neutralize the negative phospho-sugar backbone of the DNA, augmenting the DDB2-DNA binding interactions and contributing to the overall high-binding affinity for damaged DNA. (C) The DDB1-DDB2 molecular interface displays both aliphatic and charge characteristics, with the distribution of charged (in blue & red) and hydrophobic (in white) residues aligned to be complementary and neutralize the overall electrostatic potential upon complex formation.

Figure S3: Schematic representations of the multiple sites of interactions found at the abasic lesion between the dimeric DDB2 and the AP24 oligodeoxynucleotide .

(A) The dimeric DDB2, depicted by yellow ribbon representation, with an AP24 oligodeoxynucleotide mer bound at each site in the monomers. The NT region of DDB2 folds into a helical bundle ("α-paddle") that forms extensive interactions with a neighboring DNA molecule, depicted by stick-and-ball representation; the phospho-ribose backbone is colored in orange-red for the damaged strand and in blue for the undamaged strand. Defined sites of interactions between DDB2 and the damaged DNA can modulate binding affinities and optimize positioning of the damaged and undamaged DNA strands. Four distinct groups of interactions are found at the primary DNA binding site in DDB2 and mapped to the DDB2 structure, with a sphere representing the position of the DDB2 residue. At the lesion site on the damaged strand (D+1), residues at the β-loop are represented by an orange sphere; residues that stabilize the THF11 and dC12 (D+1) in their extrahelical configuration are denoted by yellow spheres; residues forming predominantly electrostatic interactions with the phosphodeoxyribose backbone of the DNA within a two basepair window upstream and downstream of the lesion sites are

denoted by violet (D-1,D-2) and green (D+3,D+4) spheres. (B**)** The complete 24 base pair duplex containing an apurinic lesion mimicked at position 11 (THF11), with the same color scheme representing the four groups of DDB2-DNA interactions with the damaged DNA strand. Additionally, DDB2 residues that form interactions with the undamaged DNA strand are represented by tan rectangles, with the dimer junction denoted by Asn360, colored in red. **(**C) Additional DDB2-DNA interactions are found in the dimeric DDB2; the residues from one monomer subunit are denoted by ovals while residues from the other subunit are denoted as rectangles, using the same color scheme as in Figures 2A and 2B. DDB2 interactions with the undamaged strand are predominantly electrostatic in nature. A large cluster of lysines and arginines is found opposite the lesion site and bridges between the two strands. These charge clusters may denote poly-ubiquitination sites, especially as DDB2-DNA contacts are likely to be disrupted with a cascade effect, resulting in cooperative dissociation, based on the alignment of these charge residues.

Figure S4: Evidence for UV-DDB monomer to dimer transition when bound to damaged DNA as revealed by volume analysis using AFM.

A. Representative surface plot of UV-DDB (50 nM) in the presence of undamaged 517 bp PCR fragments (25 nM) showing free UV-DDB (white arrow) and bound to the undamaged DNA (yellow arrow). The image is at 500 nm by 500 nm and 3 nm height scale. **B.** Volume analysis of UV-DDB bound to undamaged DNA. Inset: Gaussian fits to the data in the range 0 – 175 nm^3. Data points outside this range were ignored to obtain the Gaussian fit. From the fit, the peak positions correspond to 81 and 133 nm^3 with an R^2= 0.9 consistent with the size of the heterodimer (DDB1-DDB2) and dimer of heterodimers of UV-DDB $(DDB1\text{-}DDB2)_2$. **C.** Summary of fraction of UV-DDB bound to DNA for undamaged (gray; 37%, N=163 and UV-

irradiated (black; 96%, N=435 DNA. Fractions were calculated a (number of UV-DDB protein-DNA complexes)/ (number of UV-DDB protein-DNA complexes) + (number of free UV-DDB protein molecules on the mica). **D.** Overlay of percentage histograms for AFM derived volumes of UV-DDB molecules calculated on undamaged (dashed line) and UV-irradiated DNA (solid line). These data were generated by normalizing the count in each bin with respect to the total number of counts and converting to a percentage for each bin (undamaged DNA or UV-irradiated DNA). **E.** Histogram of the differences in the percentage histograms for UV-DDB bound to UV-irradiated DNA with respect to the undamaged DNA calculated bin wise.

Figure S5: Calibration curve for Nanoscope V.

The AFM volume of a protein (V) is a linear function of its molecular weight (MW). Calibration curve for Nanoscope V was constructed by performing AFM volume analysis of the following proteins: Pot1 (65 kDa), PcrA (86 kDa), UvrA monomer (105 kDa), Taq MutS dimer (181 kDa), UvrA dimer (210 kDa) and Taq MutS tetramer (362 kDa). The equation for the AFM volume is : V (nm^3) = 0.3856 MW (kDa) – 1.913 with R^2 = 0.9886. Based on this equation, the AFM volume of DDB1-DDB2 heterodimer (MW
= 175 kDa) is 66 nm^3 (dotted arrows) and that of a dimer of the DDB1-DDB2 heterodimer (MW = 350 kDa) is 133 nm^3 (dashed arrows).

Figure S6: XP-E mutations mapped on to the dimeric human UV-DDB structure.

(A,B). Two views showing the locations of common DDB2 mutations found in XP-E patients *(also see Table S5)*. These residues (represented by atomic Van der Waals surfaces) form multiple, bridging contacts to DDB1 (Leu350, Arg273) or DNA (Lys244, Asp307). Disruption of these interactions presumably lead to more global destabilization in UV-DDB, as these residues

appear to maintain multiple intermolecular associations in the holo-complex. (C,D). In the dimeric UV-DDB, the deleterious effects of two mutations in particular, Leu350 and Asn307, are more apparent, as these additionally modulate interactions to the NT α-paddle domain of DDB2. Thus, Leu350 and Asp307 of DDB2 form key intermolecular associations, at the junctions to both DDB1 and to the damaged DNA.

Figure S7: Representative electron density map of the dimeric UV-DDB-AP24 complex, **using combined SeMet anomalous phasing approaches combined with partial model phasing methods.**

Representative electron density map, using a combined SeMet anomalous phasing approach along with partial model phasing methods. The initial 2Fo-Fc map, contoured at 1.8σ, clearly showing the AP24 oligodeoxynucleotide bound at one surface of the β−propeller of DDB2.

Figure S8: Dynamic Light Scattering (DLS) analysis showing the particle sizes of UV-DDB and composition (% mass), determined at various protein complex concentrations and in the absence & presence of damaged DNA (AP24).

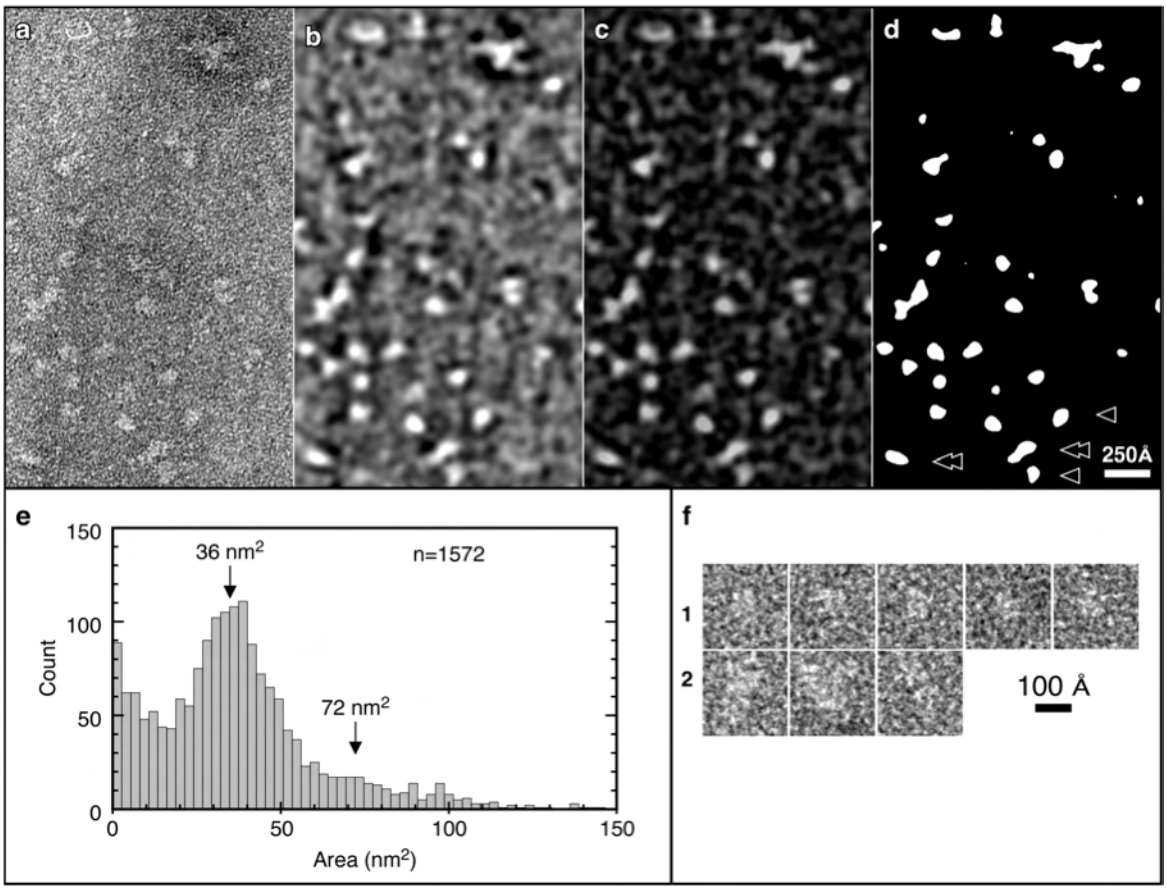
a
b
c
d
250Å
e
150
100
50
0
Count
36 nm²
n=1572
72 nm²
0
50
100
150
Area (nm²)
f
1
2
100 Å

Figure S1

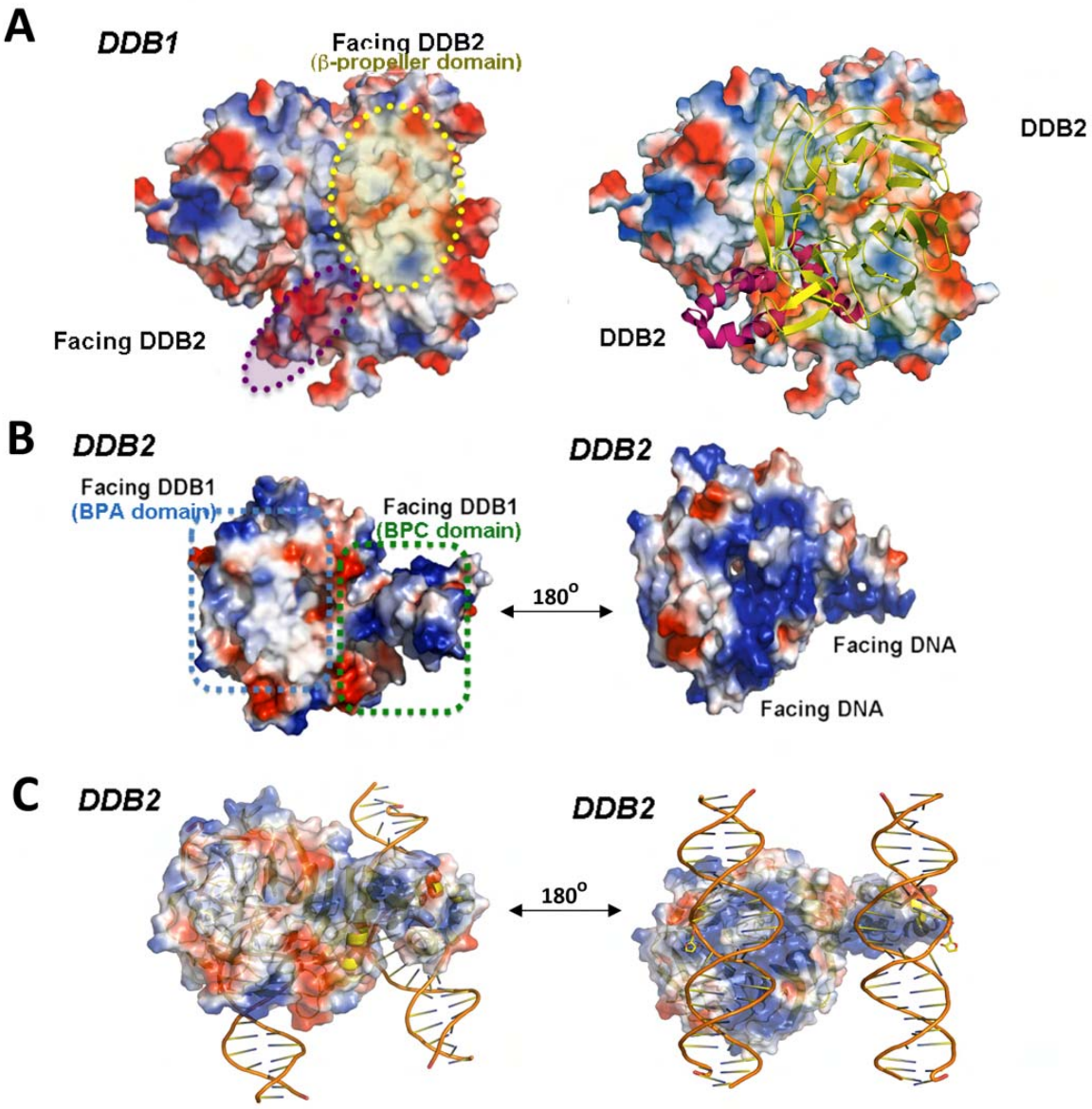
A
DDB1
Facing DDB2
(β-propeller domain)
Facing DDB2
DDB2
DDB2
B
DDB2
Facing DDB1
(BPA domain)
Facing DDB1
(BPC domain)
180°
DDB2
Facing DNA
Facing DNA
C
DDB2
180°
DDB2

Figure S2

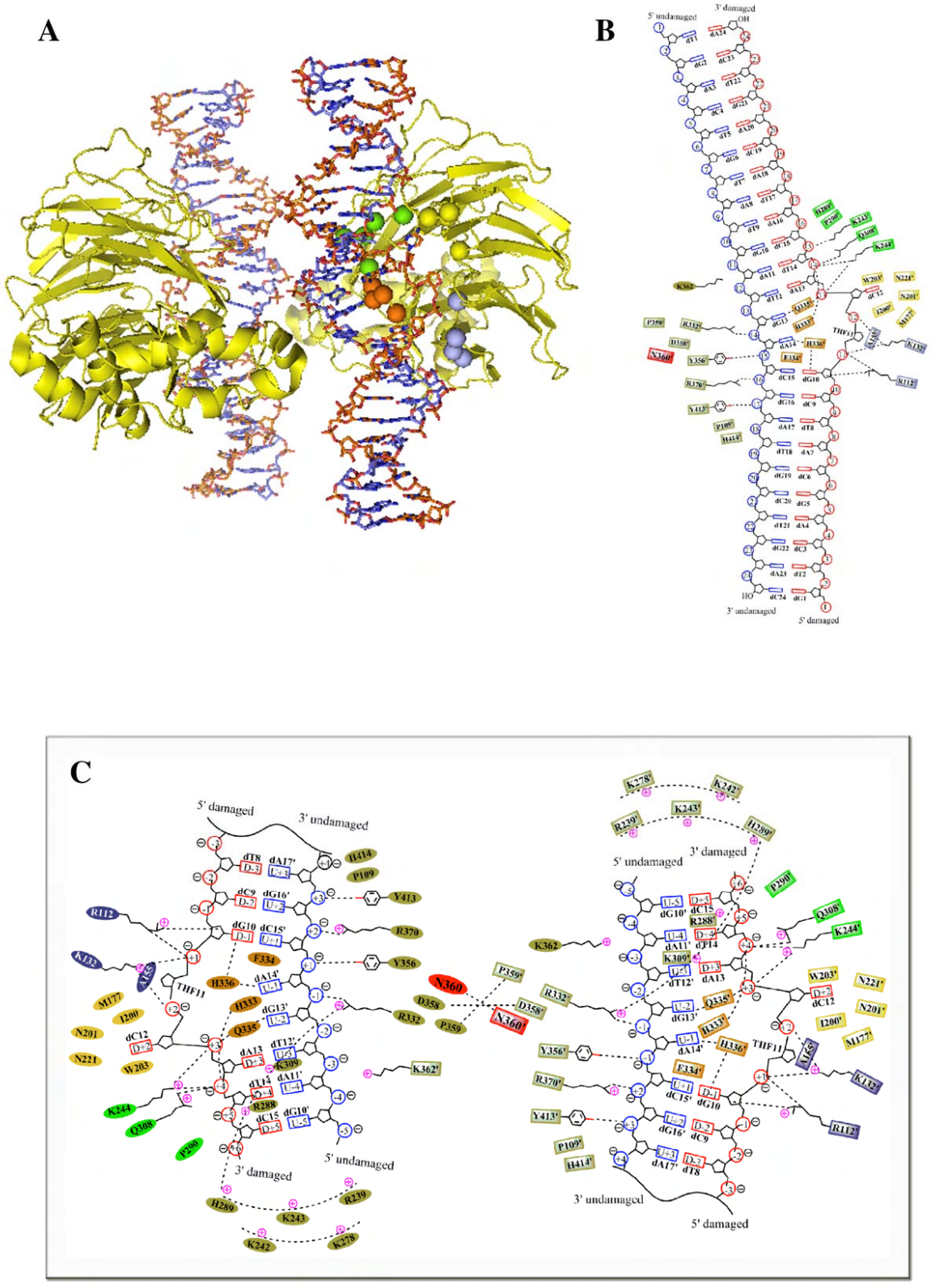

Figure S3

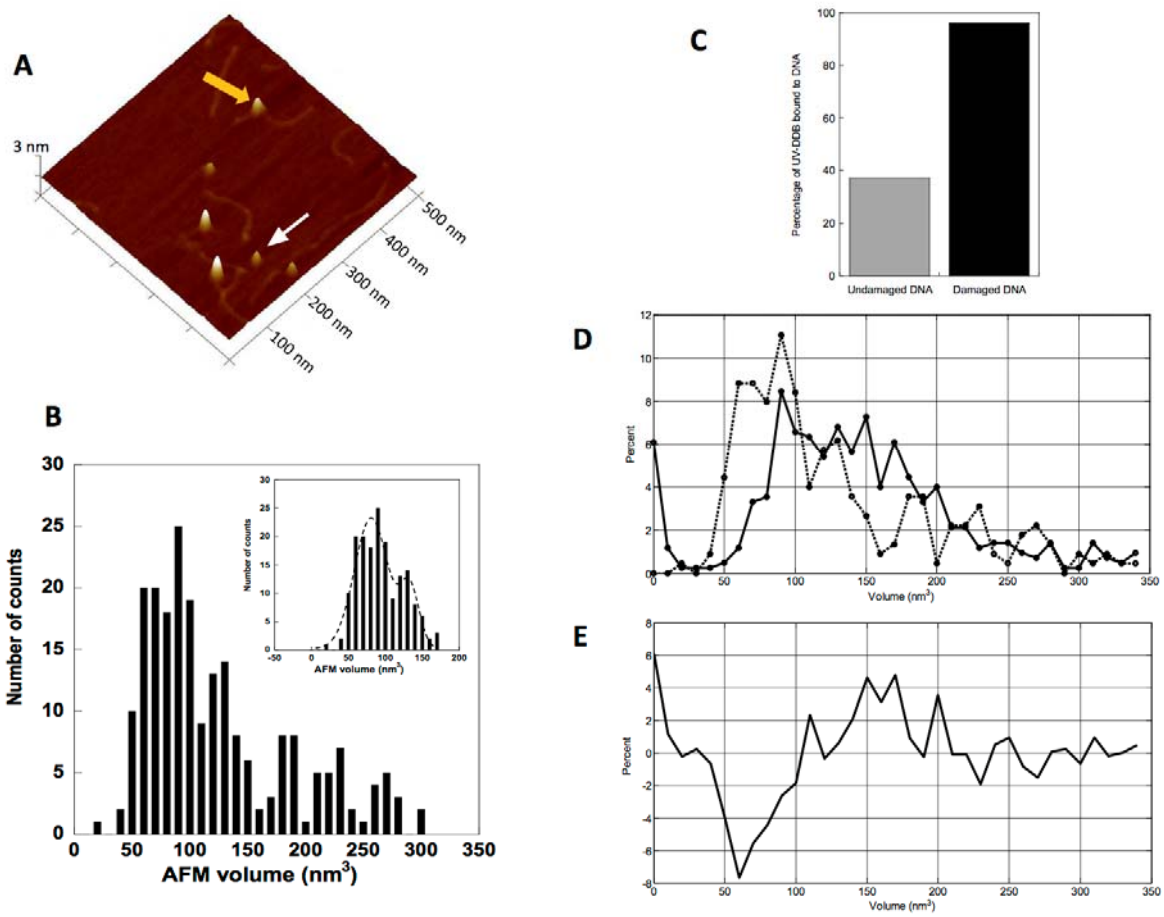
A
3 nm
500 nm
400 nm
300 nm
200 nm
100 nm
B
Number of counts
AFM volume (nm³)
C
Percentage of UV-DDB bound to DNA
Undamaged DNA
Damaged DNA
D
Percent
Volume (nm³)
E
Percent
Volume (nm³)

Figure S4

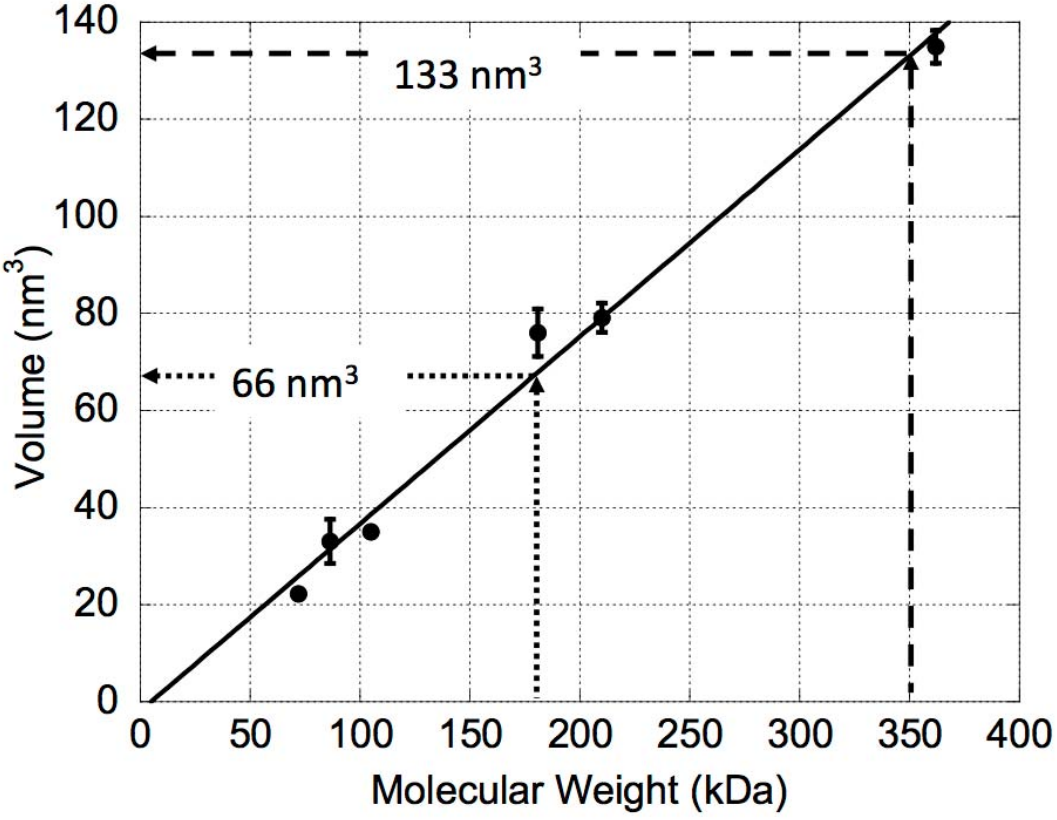
140
120
100
80
60
40
20
0
Volume (nm³)
133 nm³
66 nm³
0
50
100
150
200
250
300
350
400
Molecular Weight (kDa)

Figure S5

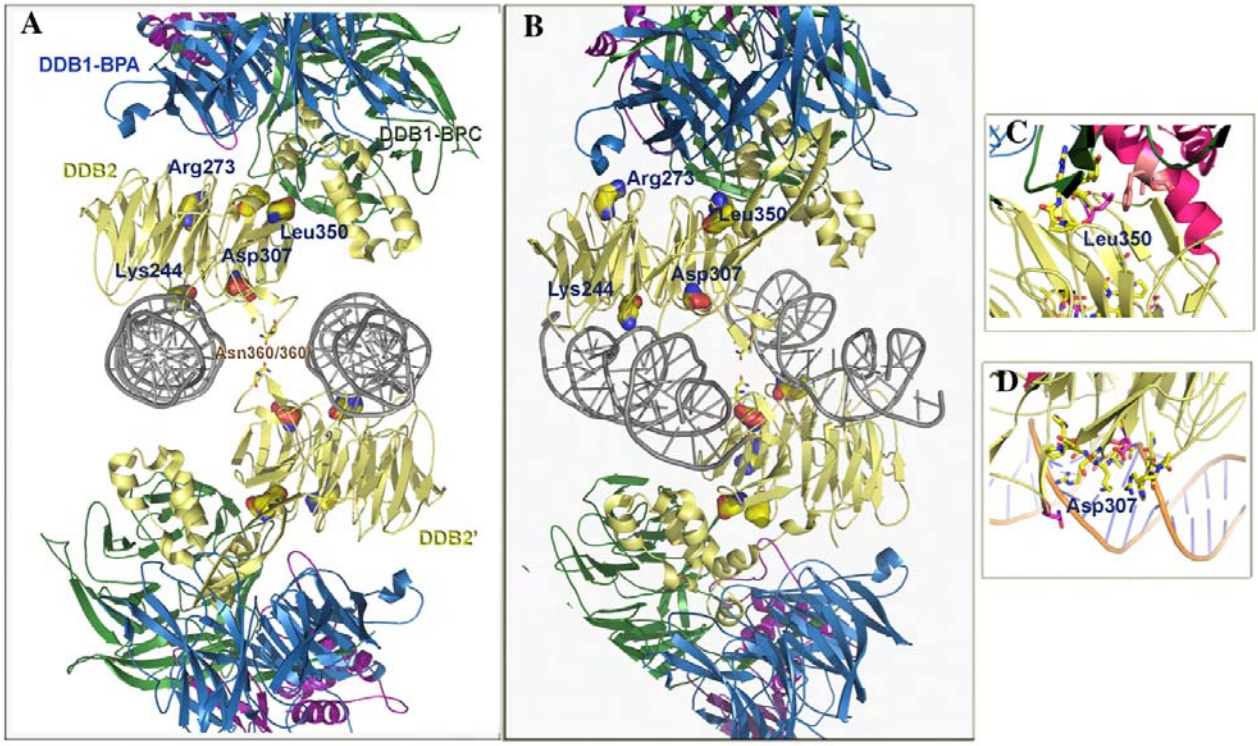
A
DDB1-BPA
DDB1-BPC
DDB2
Arg273
Leu350
Lys244
Asp307
Asn360/360'
DDB2'
B
Arg273
Leu350
Lys244
Asp307
C
Leu350
D
Asp307

Figure S6

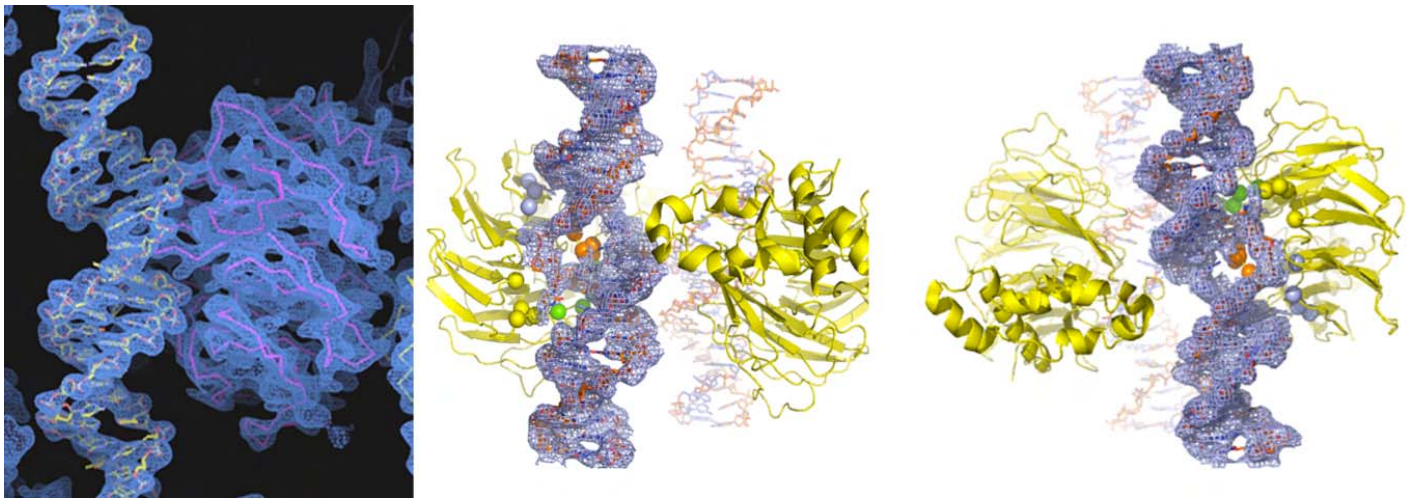

Figure S7

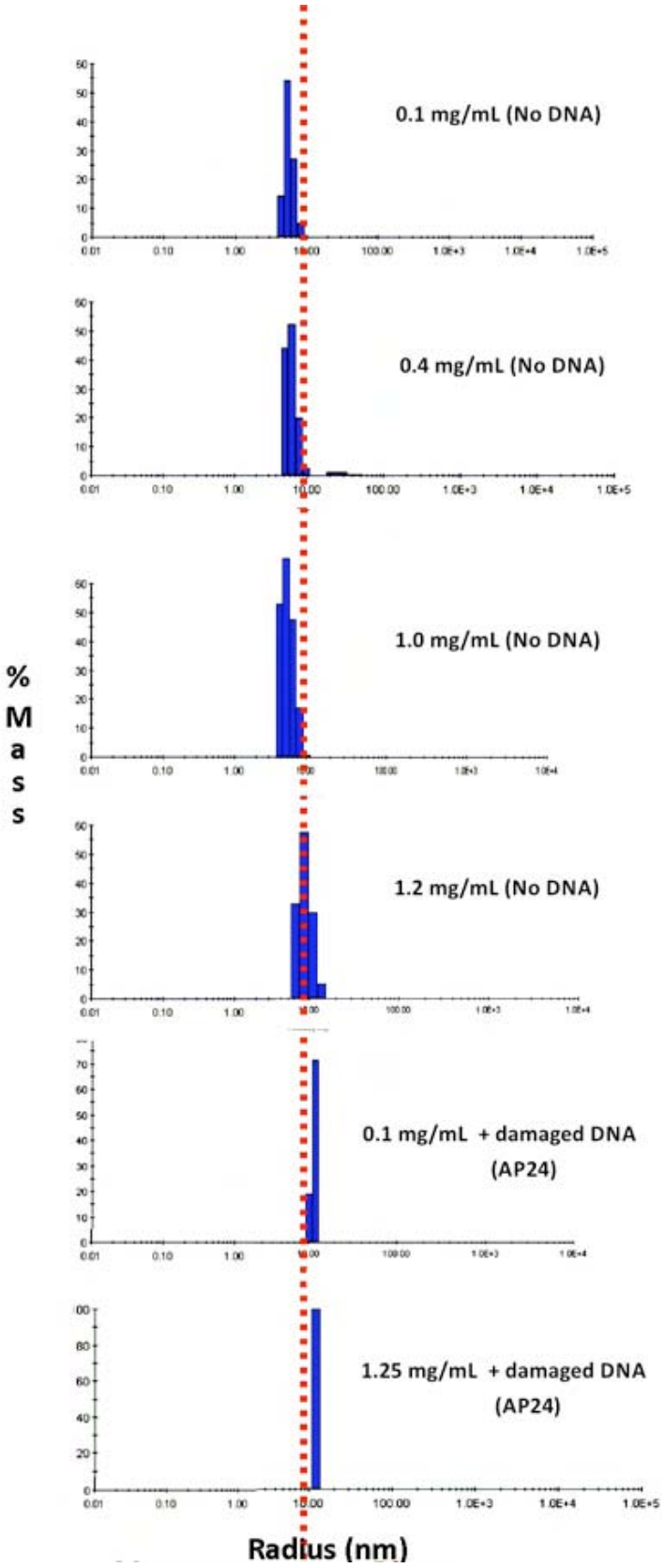
0.1 mg/mL (No DNA)
0.4 mg/mL (No DNA)
1.0 mg/mL (No DNA)
1.2 mg/mL (No DNA)
0.1 mg/mL + damaged DNA (AP24)
1.25 mg/mL + damaged DNA (AP24)
% Mass
Radius (nm)

Figure S8