Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of the C-terminal half of human XPB helicase and the impact of the disease-causing mutation XP11BE

Eduardo Hilario, Yang Li, Yumiko Nobumori, Xuan Liu and Li Fan*

Department of Biochemistry, University of California, Riverside, Riverside, CA 92521, USA

Correspondence e-mail: lifan@ucr.edu

Received 23 July 2012
Accepted 30 October 2012

PDB Reference: C-terminal half of human XPB helicase, 4ern

XPB is a DNA-dependent helicase and a subunit of the TFIIH complex required for both transcription and DNA repair. XPB contains four domains: an N-terminal domain, two conserved helicase domains (HD1 and HD2) and a C-terminal extension. The C-terminal extension is important for DNA repair since the phosphorylation of Ser751 inhibits 5′-incision by ERCC1-XPF endonuclease. A disease-causing frameshift mutation (XP11BE) that changes the last 42 amino acids of XPB causes manifestations including impaired DNA repair and deficient transcription. Here, the crystal structure of the C-terminal half of XPB (residues 494–782) is reported at 1.8 Å resolution. The structure contained the conserved XPB HD2 and a C-terminal extension which shares structural similarity with RIG-I, leading to a structural model of the XPF–XPB–DNA complex for 5′ incision during DNA repair. A mutation mimicking the XP11BE mutation produced the much less soluble mutant XPBm(494–781). Western blotting results confirmed that the intracellular levels of XPB and other TFIIH subunits in XP11BE patient cells were much lower than those from the healthy parents. Together, these results indicate that the XP11BE mutation not only divests the XPF-interaction motif, impairing DNA repair, but also reduces XPB solubility, leading to a lower intracellular level of TFIIH and deficient transcription.

1. Introduction

The human xeroderma pigmentosum (XP) group B (XPB) gene (also called ERCC3) encodes a 3′–5′ DNA helicase that is a component of the general transcription factor TFIIH complex (Schaeffer *et al.*, 1993). TFIIH is required for both basal and activated transcription as well as nucleotide-excision repair (NER), a major DNA-repair pathway that fixes UV-light-induced photoproducts and various DNA helix-distorting lesions caused by chemicals (for reviews, see Bootsma *et al.*, 2002; Lehmann, 2003; Gillet & Schärer, 2006; Wood, 1997). NER involves over 30 proteins that mediate DNA repair through a 'cut-and-patch' mechanism *via* two sub-pathways: transcription-coupled repair (TCR) and global genome repair (GG-NER). TCR and GG-NER differ in the initial damage-recognition step: DNA damage causes RNA polymerase II to stall during transcription elongation in TCR, whereas a damage-recognition complex detects the lesion through 'scanning' the genome in GG-NER. After damage recognition, the two sub-pathways share a common process involving the recruitment of TFIIH, which unwinds the double-stranded helical DNA around the lesion mediated by XPB and XPD, a 5′–3′ DNA helicase within the TFIIH complex (Fan *et al.*, 2008). The resulting DNA bubble is recognized by XPG and ERCC1-XPF, two junction-specific

endonucleases with opposite single-strand polarities (O'Donovan *et al.*, 1994; Sijbers *et al.*, 1996). XPG and ERCC1-XPF generate dual incisions on the same strand at the 3′ and 5′ sides of the lesion, respectively, resulting in the removal of a 24–32-base-long single-stranded DNA containing the lesion. The gap is finally filled by DNA polymerase and ligase (Shivji *et al.*, 1995) using the undamaged DNA strand as the template.

The importance of XPB helicase is attested by its disease-causing mutations. Genetic mutations in XPB are associated with three diseases: XP, combined XP and Cockayne syndrome (XP/CS), and trichothiodystrophy (TTD) (Bootsma *et al.*, 2002; Oh *et al.*, 2006; Hoeijmakers, 1994; Cleaver *et al.*, 1999; Kraemer, 2003). XP patients display abnormal sensitivity to sunlight and an increased risk of skin cancer owing to NER deficiency (Oh *et al.*, 2006). Patients with the XP/CS complex have the features of XP combined with the neurological abnormalities of CS (Weeda *et al.*, 1990; Scott *et al.*, 1993; Vermeulen *et al.*, 1994; Rapin *et al.*, 2000). TTD patients have neurological abnormalities and characteristically brittle hair, but no increased frequency of skin cancer (Weeda *et al.*, 1997; Riou *et al.*, 1999). Both CS and TTD have characteristics of defective transcription, reflecting the important role of XPB in transcription. The transcription of protein-coding genes is mediated by RNA polymerase II (Pol II) and starts with the sequential recruitment of general transcription factors at the promoter of the transcribed gene to form the pre-initiation complex (PIC), which contains Pol II, TFIIH and other transcription factors (Compe & Egly, 2012). After PIC formation, the ATPase activity of XPB is required for promoter opening (Dvir *et al.*, 1996; Holstege *et al.*, 1996; Kim *et al.*, 2000; Douziech *et al.*, 2000) and escape (Moreland *et al.*, 1999) to initiate the synthesis of RNA by Pol II.

The first XP-B patient, designated XP11BE (Noojin, 1965; Robbins *et al.*, 1974; Brumback *et al.*, 1978), suffered from a severe XP/CS phenotype characterized by XP features (extreme sensitivity to sunlight with blistering in infancy, pigmentation abnormalities and multiple skin cancers at an early age) and the CS phenotype (wizened facial appearance, dwarfism, sensorineural deafness, microcephaly, severe mental retardation and immature sexual development). Genetic analysis revealed a mutation in the XPB gene that caused a C–A transversion in the last intron, generating a 4 bp insertion in the mRNA and a frameshift that produces an XPB mutant consisting of only 781 amino acids and changes the sequence of the last 41 residues (Weeda *et al.*, 1990). This mutation is likely to impair interactions between XPB and ERCC1-XPF, as phosphorylation of Ser751 inhibits 5′-incision by ERCC1-XPF during DNA repair (Coin *et al.*, 2004). In addition, *in vivo* genetic analysis with yeast (Guzder *et al.*, 1994) has shown that deletion of the sequences equivalent to the last 42 residues of human XPB in the yeast homolog Rad25 impairs DNA repair. However, this deletion has no effect on transcription. Furthermore, mouse models (Andressoo *et al.*, 2009) with a similar deletion in XPB displayed the XP phenotype but no CS manifestation. It remains a mystery as to why the XP11BE patient developed the severe XP/CS complex rather than the XP phenotype. Here, we report our structural and biochemical characterizations of the C-terminal half of XPB (XPB-C; residues 494–782) and provide insights into the impact of the disease-causing mutation XP11BE.

2. Materials and methods

2.1. Cloning, expression and purification of recombinant XPB-C protein

The cDNA sequence encoding amino-acid residues 494–782 of human XPB was amplified by polymerase chain reaction (PCR) using Phusion High-Fidelity DNA Polymerase (NEB BioLabs, USA) with the plasmid pOTB7-HsXPB (The CCSB Human ORFEOME collection) as the template (see Supplementary Material[1] for details). The final PCR products corresponding to DNA encoding XPB-C and XPBm(494–781) were cloned separately into the *Escherichia coli* expression vector pGEX-6P1 (GE Healthcare, USA). Protein expression was under the control of the isopropyl β-D-1-thiogalactopyranoside (IPTG) inducible T7 promoter in *E. coli* Rosetta (DE3) pLys-S cells (Invitrogen, USA). Protein expression was induced with 0.1 m*M* IPTG overnight at 298 K. Cells from 6 × 1 l culture were harvested and lysed with 300 ml buffer *A* (2× PBS pH 7.5 containing 5% glycerol, 0.5% Nonidet NP-40, 2 m*M* EDTA, 2 m*M* EGTA, 10 m*M* benzamidine, 1 m*M* PMSF, 1 m*M* β-glycerophosphate, 1 m*M* sodium pyrophosphate and 1 m*M* sodium vanadate) by sonication (Branson Sonifier D450). The supernatant fraction was clarified at 50 000*g* for 20 min at 277 K and loaded onto a 5 ml Glutathione Sepharose column (GE Healthcare) using an ÄKTA-purifier UPC10 (GE Healthcare) at 281 K. Recombinant GST-tagged protein was eluted with buffer *B* (50 m*M* Tris–HCl pH 8.0, 10 m*M* reduced glutathione, 100 m*M* NaCl, 5% glycerol, 1 m*M* EDTA, 10 m*M* benzamidine, 1 m*M* PMSF, 1 m*M* β-glycerophosphate, 1 m*M* sodium pyrophosphate, 1 m*M* sodium vanadate). Aliquots of the eluate fractions were analyzed by 12%(*w*/*v*) SDS–PAGE and fractions containing the GST-XPB-C protein were pooled and concentrated using Amicon Ultra 30K filters (Millipore, USA) at 277 K and 3000*g*. The concentrated sample was diluted in PreScission protease buffer (GE Healthcare, USA) and digestion was carried out according to the manufacturer's recommendations. The digestion reaction mixture was passed through a GSTrap FF 5 ml column and the flowthrough fraction containing the XPB-C protein was collected and concentrated using Amicon Ultra 10K filters (Millipore, USA) at 277 K. The concentrated sample was further purified using a HiPrep 16/60 Sephacryl S-100 High Resolution size-exclusion column (GE Healthcare) in buffer *C* (10 m*M* Tris–HCl pH 8.0 containing 100 m*M* NaCl, 5% glycerol). Samples corresponding to the peak fractions were analyzed by 15%(*w*/*v*) SDS–PAGE. Fractions containing pure HsXPB-C were pooled, concentrated using Amicon Ultra 10K filters and stored in aliquots containing protein at 10 mg ml^{-1} at 193 K for crystal preparation.

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: MN5013). Services for accessing this material are described at the back of the journal.

2.2. Crystallization, data collection and data processing

Initial crystallization screening was performed at the Macromolecular X-ray Core Facility at UCR College of Natural and Agricultural Sciences. Briefly, sitting-drop crystallization experiments were performed in Intelli-Plate 96-3 plates using a Phoenix robotic crystallization system (Art Robbins Instruments, Sunnyvale, California, USA) by mixing 200 nl protein sample at 10 mg ml^{-1} with 200 nl of the reservoir solutions from commercial crystallization kits. The plates were then stored at room temperature in the CrystalMation system and monitored using a Desktop Minstrel UV automation system (Rigaku, USA). Protein crystals were observed in condition E10 (0.1 *M* sodium citrate pH 5.6, 0.2 *M* ammonium acetate, 30% PEG 4000) of The JCSG Core II Suite (Qiagen, Valencia, California, USA). This initial condition was refined by the hanging-drop vapor-diffusion method using 24-well Linbro plates containing 500 µl reservoir solution and 4 µl drops (2 µl protein solution and 2 µl reservoir solution) at room temperature. The crystals used for data collection were obtained using a reservoir solution consisting of 0.1 *M* sodium acetate pH 5.55, 0.2 *M* ammonium acetate, 27% PEG 3350. For data collection, a single crystal was soaked in reservoir solution containing 5% glycerol as a cryoprotection agent for a few minutes and then mounted in a nylon loop (Hampton Research, USA) and flash-cooled directly in a nitrogen stream at 100 K. Diffraction data were collected at the Macromolecular X-ray Core Facility at UCR using a Rigaku MicroMax-007 HF copper rotating-anode X-ray generator (1.5418 Å wavelength) and an R-AXIS IV^{++} imaging-plate detector. Data were indexed and integrated to 1.8 Å resolution using *iMOSFLM* (Battye *et al.*, 2011) and scaled and merged with *SCALA* (Evans, 2006).

2.3. Structure determination and refinement

The structure of XPB-C was solved by the molecular-replacement method using *Phaser* (McCoy *et al.*, 2007). The C-terminal domain of AfXPB (PDB entry 2fzl; Fan *et al.*, 2006) was used as the search model after the solvent molecules had been removed and the residues had been mutated based on the result of sequence alignment (Fan *et al.*, 2006) to generate a new model prior to molecular replacement. The initial molecular-replacement result was refined using rigid-body refinement in *REFMAC* (Murshudov *et al.*, 2011), followed by automatic model building with *ARP*/*wARP* (Perrakis *et al.*, 1999). There was one XPB molecule in each asymmetric unit. The model was then manually improved in *Coot* (Emsley *et al.*, 2010) and refinement was performed with *PHENIX* (Adams *et al.*, 2010). Several cycles of alternate manual building and refinement led to the final model (Table 1). All protein-structure molecular graphics in the figures were prepared with *PyMOL* (DeLano, 2002).

2.4. Structure modeling

The structural models shown in Figs. 3–5 were manually built with *PyMOL* (DeLano, 2002) based on the alignments revealed by the *DALI* server (Holm & Rosenström, 2010).

Table 1
Statistics of data collection and refinement (molecular replacement).

Values in parentheses are for the highest resolution shell.

PDB code	4ern
X-ray source	MicroMax-007 HF
Wavelength (Å)	1.5418
Space group	$P2_12_12_1$ [No. 19]
Unit-cell parameters (Å, °)	a = 38.26, b = 73.65, c = 84.30, $\alpha = \beta = \gamma$ = 90.00
No. of reflections	105681 (14414)
No. of unique reflections	22299 (3171)
Resolution range (Å)	19.29–1.80 (1.90–1.80)
Multiplicity	4.7 (4.5)
$\langle I/\sigma(I)\rangle$	20.7 (3.0)
Completeness (%)	98.2 (97.2)
$R_{\rm merge}$† (%)	3.6 (48.8)
Refinement statistics	
Resolution (Å)	18.55–1.80
Total No. of reflections	22258
$R_{\rm work}/R_{\rm free}$‡ (%)	20.10/23.77
No. of non-H atoms	
Protein	1916
Water	86
Average B factors (Å^2)	
Protein	38.61
Water	45.09
R.m.s. deviations from ideal geometry	
Bond lengths (Å)	0.011
Bond angles (°)	1.343
Ramachandran plot	
Most favored regions	199 [94.3%]
Additionally allowed regions	12 [5.7%]
Generously allowed regions	0 [0%]
Disallowed regions	0 [0%]

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ and $\langle I(hkl)\rangle$ are the observed individual and mean intensities of a reflection with indices hkl, respectively, $\sum_i$ is the sum over i measurements of a reflection with indices hkl and $\sum_{hkl}$ is the sum over all reflections. ‡ $R = \sum_{hkl} \big||F_{\rm obs}| - |F_{\rm calc}|\big| / \sum_{hkl} |F_{\rm obs}|$. $R_{\rm free}$ is the R value calculated for 5% of the data set not included in the refinement.

The model of the XPF–XPB–DNA complex was developed by superimposing XPB-C on helicase domain 2 of RIG-I (PDB entry 2ykg; Luo *et al.*, 2011), the N-terminal half of AfXPB (PDB entry 2fz4; Fan *et al.*, 2006) on the helicase domain of RIG-I and the dsDNA of the archaeal XPB–DNA complex (PDB entry 2bgw; Newman *et al.*, 2005) on the dsRNA of the RIG-I complex (PDB entry 2ykg). The C-terminal extension of RIG-I was also reduced to 50 residues in order to match the 52-residue C-terminal tail of XPB.

2.5. Cell culture

Lymphoblasts (Patient GM02254) and XP-B cells from a patient with the XP/CS complex (Patient XP11BE-GM02252), from the mother of Patient XP11BE (Patient XPH134BE-GM01855) and from the father of Patient XP11BE (Patient XPH135BE-GM11394) were obtained from the Human Genetic Mutant Cell Repository (Camden, New Jersey, USA; http://www.ccr.coriell.org/ccr). Lymphoblast cells were grown in RPMI 1640 medium with 15% fetal bovine serum.

2.6. Western blotting

Whole-cell extracts were prepared by lysing the cells in a buffer consisting of 50 m*M* Tris–HCl pH 8.0, 120 m*M* NaCl, 0.5% NP-40, 1 m*M* dithiothreitol, 2 µg ml^{-1} aprotinin and

2 µg ml^{-1} leupeptin followed by centrifugation to separate the supernatant from debris. The protein from cellular extracts was subjected to SDS–PAGE using 10% gels, transferred to Protran nitrocellulose membrane (Whatman, Kent, England) and probed separately with the specific antibodies anti-XPB rabbit serum (prepared by Covance Research Products, Denver, Pennsylvania, USA using purified XPB C-terminal half protein), polyclonal antibodies against XPB (XPB-sc), XPD, XPF and p62 (Santa Cruz Biotechnology, Santa Cruz, California, USA), and antibody against vinculin (Sigma, St Louis, Missouri, USA). The proteins were detected using the ECL Western blot detection reagents (Millipore, Billerica, Massachusetts, USA). The secondary peroxidase-conjugate immunoglobulin G (IgG-HRP) was obtained from Bio-Rad (Hercules, California, USA). Proteins were visualized by exposing X-ray films (Denville Scientific, Metuchen, New Jersey, USA) to the membranes.

3. Results and discussion

3.1. Overall structure of XPB-C

The full-length XPB helicase consists of 782 amino-acid residues containing seven conserved helicase motifs (I, Ia and II–VI) in the middle of the polypeptide. The crystal structures of an archaeal XPB homolog with a sequence equivalent to residues 240–686 of human XPB revealed an XPB core structure consisting of four domains (Fig. 1): two helicase domains, a small damage-recognition domain (DRD) attached to the N-terminus of helicase domain 1 and a thumb-like (ThM) domain inserted into helicase domain 2 (Fan *et al.*, 2006). The ThM insert is structurally similar to the thumb domain of T4 DNA polymerase. In addition to the seven helicase motifs that are conserved in most helicase superfamily 2 members, XPB homologs have a unique RED motif containing three charged residues Arg-Glu-Asp near helicase motif III (Fan *et al.*, 2006). The RED motif and the ThM insert are specifically involved in the DNA-dependent stimulation of ATPase activity of human XPB helicase (Oksenych *et al.*, 2009). In addition, there are sequence extensions at both ends of this core in human XPB helicase (Fig. 1*a*). All disease-causing mutations identified to date are located in these two extensions (Oh *et al.*, 2006). After many trials that failed to produce soluble recombinant full-length XPB helicase protein in *E. coli*, we have successfully engineered a C-terminal fragment of human XPB consisting of residues 494–782 (named XPB-C) which contains helicase domain 2, the ThM insert and the C-terminal extension (Fig. 1*b*) in order to investigate the structural basis for the role of the C-terminus of XPB in DNA repair and transcription.

XPB-C was initially fused with GST and expressed in *E. coli* (Fig. 2*a*). After purification, GST was removed by protease digestion followed by additional chromatographic steps. The crystal structure of XPB-C was determined at 1.8 Å resolution (see Table 1 for X-ray crystallographic statistics) by molecular replacement using the C-terminal domain of the archaeal XPB homolog (AfXPB; PDB entry 2fzl; Fan *et al.*, 2006) as a structural model. The final structural model consisted of residues 502–730 containing helicase domain 2 (Figs. 1*a*–1*c*) including helicase motifs IV–VI, the ThM insert and a C-terminal extension (light pink in Figs. 1*b* and 1*c*) that was not observed in the AfXPB structure. XPB-C forms a globular domain consisting of a central sheet of seven parallel β-strands sandwiched by three α-helices on one side and two α-helices on the other (Fig. 1*b*). The C-terminal extension consists of residues 670–730 forming two α-helices connected by a long loop and a β-strand, which forms the edge strand of the seven-stranded central sheet, and an unstructured tail that possibly extends to the C-terminus. The ThM insert together with one α-helix of the C-terminal extension are located at the top of the central sheet. Although the XPB-C protein used for crystallization consisted

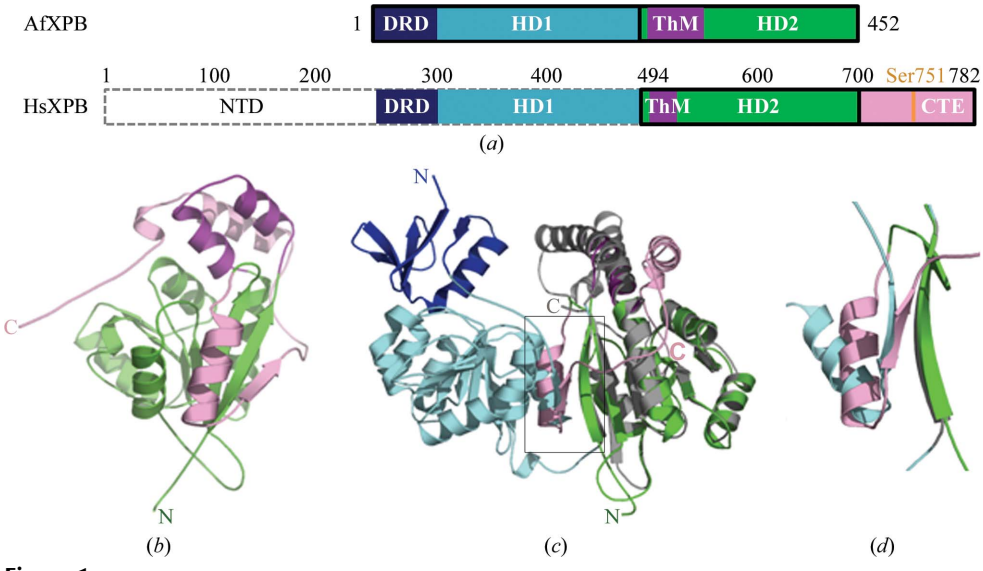


Figure 1
Structural overview of the C-terminal half of human XPB. (*a*) Schematic domain alignment between human XPB and AfXPB. Functional domains, including the N-terminal domain (NTD; white), the damage-recognition domain (DRD; blue), helicase domains 1 (HD1; cyan) and 2 (HD2; green), the thumb-like insert (ThM; magenta) and the C-terminal extension (CTE; pink), are colored differently. The known phosphate-site residue Ser751 is indicated by an orange line. The domains defined by crystal structures are shown as solid boxes. (*b*) Ribbon representation of the XPB-C structure. The functional motifs (HD2, ThM and CTE) are colored as in (*a*). (*c*) Ribbon backbones of AfXPB (constructed with PDB entries 2fwr and 2fzl) and XPB-C with the helicase domain 2 superimposed. The functional motifs are colored as in (*a*) except for the C-terminal half of AfXPB (PDB entry 2fzl), which is colored light gray to give a better view for structural comparison. The elements of XPB-C that potentially clash with the N-terminal half of AfXPB are indicated by the box. (*d*) An enlarged view of the potentially clashing elements as indicated in (*c*).

of residues 494–782, the last 52 residues (731–782) were not defined in our structure owing to a lack of electron density. Analysis of the protein in the crystals by SDS–PAGE (Fig. 2*b*) revealed that the last 52 residues were degraded during crystallization over a period of over a month at room temperature. These 52 residues either detached from the rest of the XPB-C polypeptide as a small fragment or were degraded into pieces because of an unstructured C-terminal tail. We assume that the latter is more likely because no smaller peptides apart from XPB-C were detected in the crystallization drops (Fig. 2*b*). An unfolded C-terminal tail would provide flexibility to the interaction motif located around residue Ser751 at the C-terminus of XPB so that it can adopt a proper conformation to interact with the ERCC1-XPF endonuclease during DNA repair. Many protein–protein interaction motifs are unfolded until they interact with their protein partners. For example, the PCNA-interaction peptide motif (PIP) at the C-terminus of archaeal RNase HII is unfolded (no electron density) in the structure of RNase HII at 1.95 Å resolution (PDB entry 1i39; Chapados *et al.*, 2001), but becomes ordered by forming a 3_{10}-helix and a short β-strand in the structure of the RNase HII–PCNA complex at 3 Å resolution (PDB entry 3p87; Bubeck *et al.*, 2011).

3.2. Structural comparison of XPB-C and AfXPB

There are two major differences between the structure of XPB-C and that of AfXPB (Fig. 1). Firstly, XPB-C has a smaller ThM insert consisting of residues 514–537 forming two α-helices and a short loop (Fig. 1*b*), while the ThM insert of AfXPB consists of two long α-helices connected by a long loop with a short α-helix (Fig. 1*c*). Deletion of residues 516–526, which removes the first short α-helix in the thumb-like motif of human XPB, impaired recruitment of the TFIIH complex to the damage site during DNA repair (Oksenych *et al.*, 2009), supporting the structure-based prediction that the ThM motif is likely to be involved in DNA interaction (Fan *et al.*, 2006). Secondly, the XPB-C structure contains an extension that does not exist in AfXPB (Fig. 1*c*). Interestingly, the only β-strand in this C-terminal extension is in the same position as, but in the opposite direction to, the β-strand at the C-terminal end of the long loop connecting the N-terminal DRD and the helicase domain 1 of the AfXPB homolog (Figs. 1*c* and 1*d*). In addition, the first α-helix of the XPB C-terminal extension is at a position aligned with the first α-helix of the AfXPB helicase domain 1. Considering that human XPB has a similar helicase domain 1 to that of AfXPB, we expect to observe a different orientation of helicase domains 1 and 2 in the human XPB helicase in order to avoid these two clashes shown in Figs. 1(*c*) and 1(*d*) (see below for further discussion).

3.3. Possible model of XPB interacting with dsDNA and XPF

Structural analysis using the *DALI* server (Holm & Rosenström, 2010) indicates that XPB-C shares structural similarity with many proteins that have RecA-like domains. As expected, the structure of AfXPB is at the top of this list, with a *Z*-score of 16.7. The r.m.s.d. for the alignment of 154 residues (27% sequence identity) between XPB-C and the C-terminal half of AfXPB (PDB entry 2fzl) is 1.7 Å. However, the structure-based alignment of XPB-C with most of these proteins is limited to the RecA-like domain, except for the yeast DEAD-box RNA helicase Mss116p and retinoic acid-inducible gene I (RIG-I) protein (PDB entries 4a36 and 2ykg; Kowalinski *et al.*, 2011; Luo *et al.*, 2011).

Mss116p is an ATP-dependent RNA helicase that acts as an RNA chaperone that unwinds dsRNA to facilitate the folding and splicing of mitochondrial group I and II introns (Huang *et al.*, 2005). Like human XPB, Mss166p contains an N-terminal domain, two helicase domains and a C-terminal extension with an unfolded C-terminal tail (Mohr *et al.*, 2011; Mallam *et al.*, 2012). Several crystal structures are available, including the structure of a single-stranded RNA in complex with Mss116p containing the two helicase domains and the C-terminal extension (PDB entry 3sqx; Mohr *et al.*, 2011) and the structure of a DNA/RNA duplex in complex with the helicase domain 2 and the C-terminal extension (PDB entry 4db2; Mallam *et al.*, 2012). The Mss116p helicase domain 2 and C-terminal extension share high structural similarity with XPB-C, with a *Z*-score of 14.9 and an r.m.s.d. of 3.4 Å for struc-

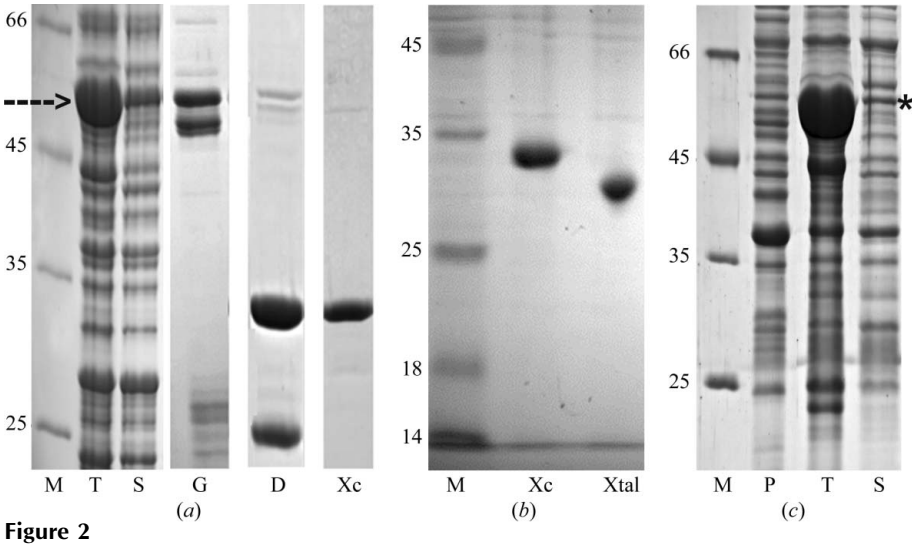


Figure 2
Comparison of the solubility of GST-XPBm(494–781) and GST-XPB-C. (*a*) SDS–PAGE results for GST-XPB-C expressed in *E. coli* and purification. (*b*) SDS–PAGE results for XPB-C protein before and after crystallization. (*c*) SDS–PAGE results for GST-XPBm(494–781) expressed in *E. coli*. Lane labels: M, EZ-Run Protein Marker (Fisher BioReagents, USA); T, cellular lysate sample (40 µg); S, soluble protein fraction after centrifugation (15 µg); G, elution fraction from GST affinity chromatography (10 µg); D, protein sample after PreScission protease cleavage (10 µg); Xc, purified XPB-C protein after S-200 gel-filtration chromatography (5 µg); Xtal, XPB-C protein from crystals; P, cellular lysate sample without expression of XPBm(494–781) (30 µg).

tural alignment over 174 residues despite a low sequence identity of 16% (*DALI* server results and Fig. 3*a*). The C-terminal extension of Mss116p including residues 506–597 extends from helicase domain 2 with a fragment consisting of a helix–loop–strand–helix–loop–helix, which is structurally very similar to the C-terminal extension observed in XPB-C (Fig. 3*b*), followed by a helix bundle which is located in a position similar to the ThM insert of XPB (Fig. 3*a*). The relative orientation of helicase domains 1 and 2 is different in Mss116p compared with that in AfXPB (Fig. 3*c*). Since the two helicase domains of human XPB cannot have the same orientation as that in AfXPB without clashes (Fig. 1*c*), we propose that the two helicase domains of XPB resemble the orientation in Mss116p (Fig. 3*d*) and have the dsDNA bound at the top groove between the two helicase domains (Fig. 3*e*). In this conformation, the dsDNA is positioned to interact with the ThM insert, which is in agreement with the results obtained by mutagenesis of the ThM motif (Oksenych *et al.*, 2009). However, it is worth noting that the C-terminal tail of XPB-C is too close to the DNA duplex. The direction of this

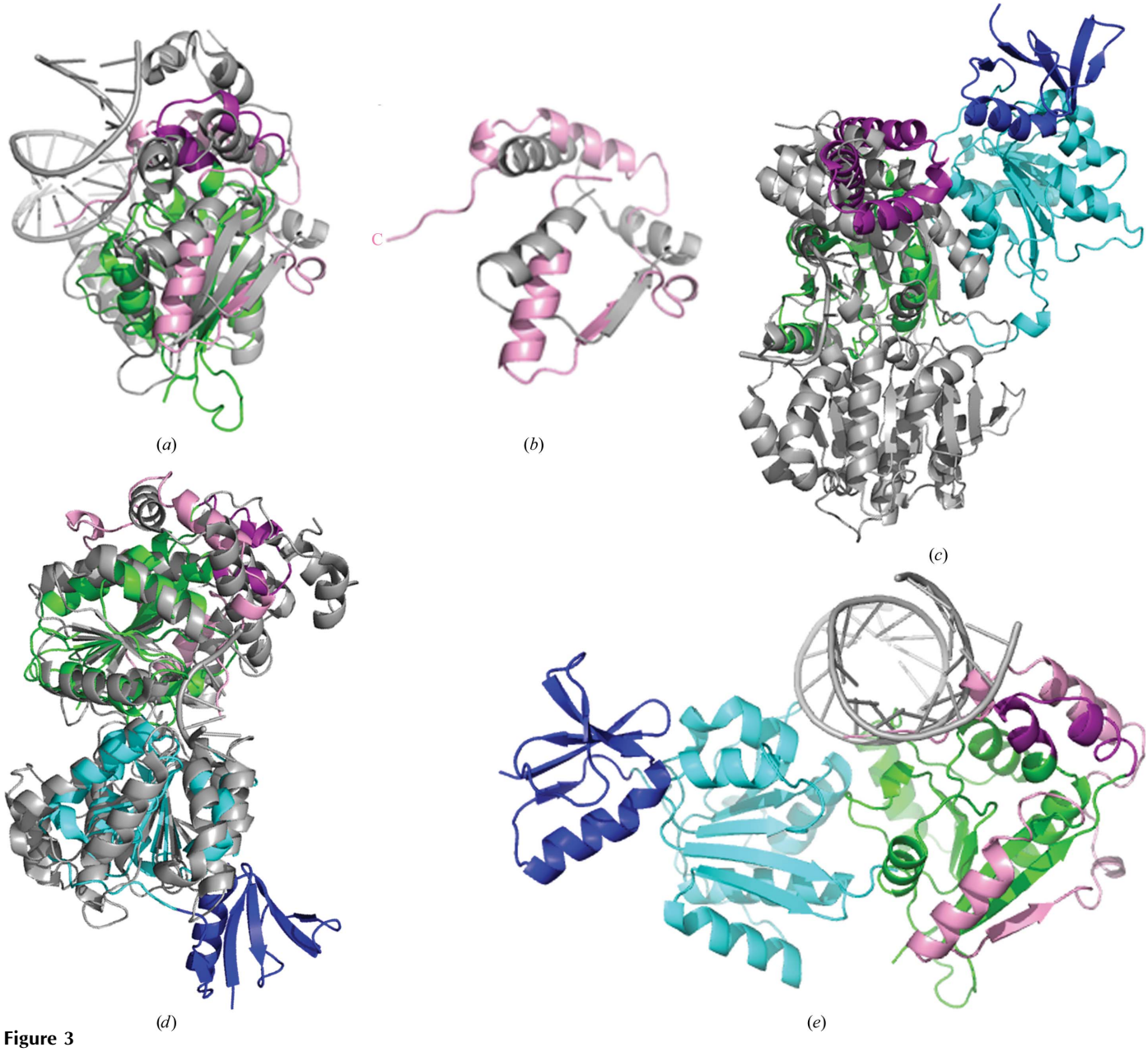
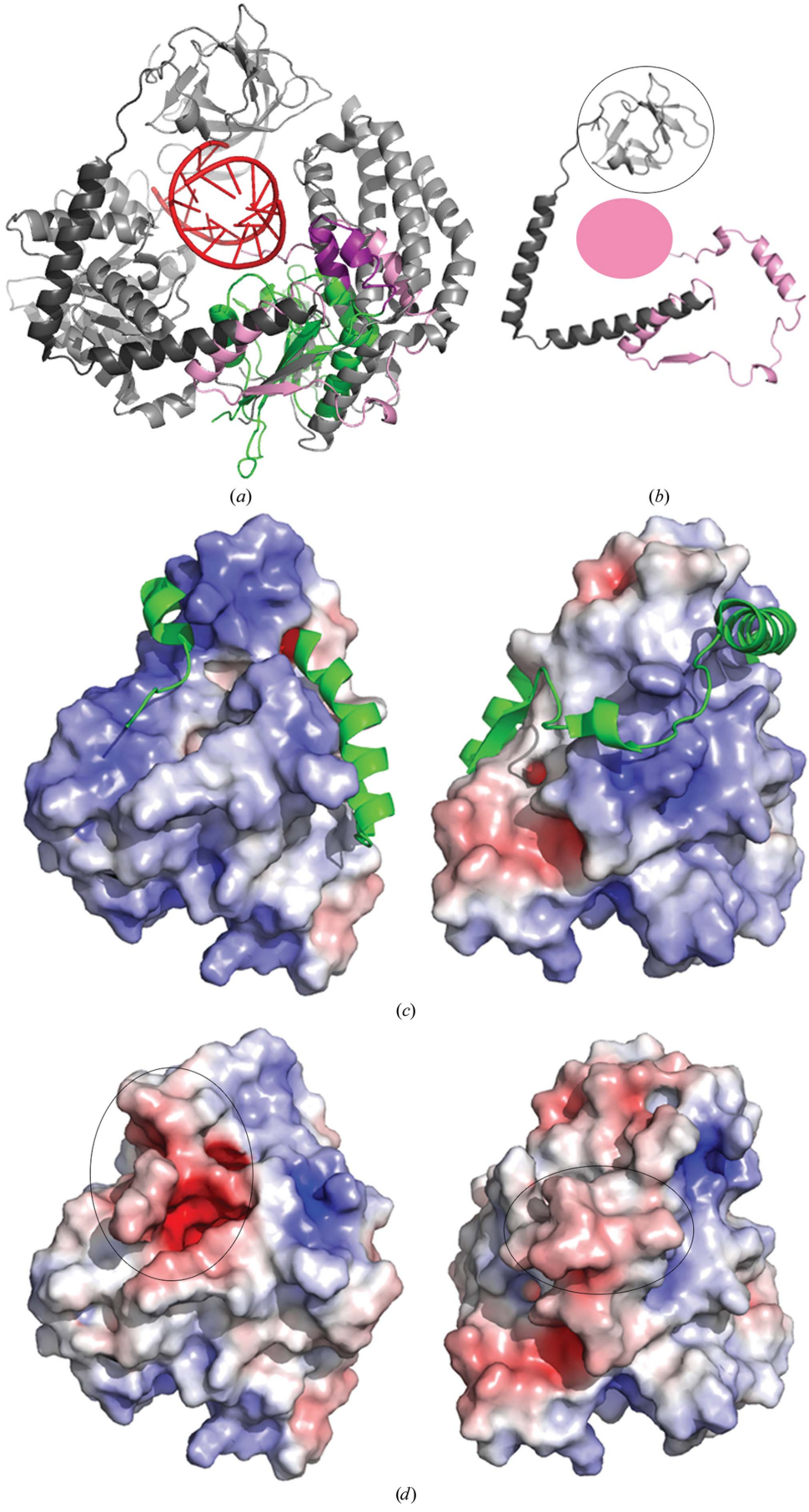


Figure 3
Structural comparison of XPB-C and yeast Mss116p. (*a*) Ribbon representation of XPB-C superimposed with helicase domain 2 of the Mss116p–dsDNA/RNA complex (light gray; PDB entry 4db2). (*b*) Close-up view of the C-terminal extension of XPB-C over the corresponding region of Mss116p shown in (*a*). (*c*) Different orientations of the helicase domains in AfXPB and Mss116p. The full-length AfXPB structure (PDB entry 2fwr) is superimposed with Mss116p (PDB entry 3qsx) over helicase domain 2. (*d*) A structural model of XPB resembling the structure of Mss116p was constructed by superimposing XPB-C and the N-terminal half of AfXPB (PDB entry 2fzl) over helicase domains 2 and 1 of Mss116p, respectively. (*e*) A structural model of the XPB–DNA complex. The model was developed by combining (*a*) and (*d*). In all panels Mss116p is displayed in gray, while AfXPB and XPB-C are colored as in Fig. 1.



tail should be modified, in particular considering that it has an extra 52 residues that are attached but are not observed in the XPB-C structure.

RIG-I is another protein identified by the *DALI* server that contains a fragment structurally similar to the C-terminal extension of XPB-C in addition to the RecA-like domain. Overall, the RIG-I structure (PDB entry 2ykg) has 166 residues that align well with those of XPB-C (Fig. 4*a*), resulting in an r.m.s.d. of 2.8 Å and a *DALI* *Z*-score of 14.4 despite only 12% sequence identity. RIG-I is a key innate immune pattern-recognition receptor that senses viral RNAs in the cellular environment and activates extensive host immunological responses against viral infection (Yoneyama *et al.*, 2004, 2005). RIG-I consists of four domains: an N-terminal domain, two helicase domains and a C-terminal domain. The N-terminal domain contains two tandem caspase activation and recruitment domains. There is also an insertion of a helix bundle in helicase domain 2. The interaction of RNA with RIG-I is

Figure 4
Structural comparison of XPB-C and RIG-I. (*a*) Ribbon diagrams of XPB-C superimposed with helicase domain 2 of RIG-I (PDB entry 2ykg) based on *DALI* alignment. The ribbons of XPB-C are colored as in Fig. 1. The ribbons of RIG-I are colored gray with the 'bridge' in dark gray. (*b*) Close-up view of the RIG-I 'bridge' superimposed on the CTE of XPB-C. The C-terminal domain of RIG-I is circled, while the missing C-terminal tail of XPB is highlighted by a solid pink ellipse. (*c*) Front (left) and back (right) views of the solvent-accessible electrostatic potential surfaces of XPB-C without the CTE included in the calculation. The CTE is shown as green ribbons on the surface. (*d*) Front (left) and back (right) views of the solvent-accessible electrostatic potential surfaces of XPB-C. The negatively charged areas caused by the CTE are circled. Electrostatic potentials were calculated using the *APBS* (http://agave.wustl.edu/apbs) *PyMOL* plugin.

mainly mediated by the two helicase domains and the C-terminal domain (Luo *et al.*, 2011; Kowalinski *et al.*, 2011). The C-terminal domain is connected to helicase domain 2 by a helix–loop–helix motif (named the bridge or pincer), forming an 'elbow'-like structure (Luo *et al.*, 2011; Kowalinski *et al.*, 2011; Fig. 4*a*). This 'bridge' of RIG-I is aligned with the C-terminal extension of XPB-C by the *DALI* server (Fig. 4*b*). The C-terminal extension of XPB-C, like the 'bridge' of RIG-I, connects helicase domain 2 and a C-terminal tail (the solid pink oval shape in Fig. 4*b*), which is not defined in our XPB-C structure but is likely to form a certain structure when XPB interacts with other partners such as XPF, as discussed above. It is possible that the C-terminal extension forms a similar 'elbow'-like structure in the full-length XPB structure. However, owing to the lack of helicase domain 1 to support the 'elbow'-like structure in XPB-C (Fig. 4*a*), the C-terminal extension is forced to fold back to wrap around helicase domain 2 as observed in the XPB-C structure. As shown in Fig. 4(*c*), the electrostatic potential surface of helicase domain 2 (including the ThM insert) is predominantly positive without the C-terminal extension, while the electrostatic potential surface of the XPB-C structure is balanced by significant negative areas on the addition of the C-terminal extension (Fig. 4*d*). These results indicate that the C-terminal extension is negatively charged and that its interactions with the positively charged surface of the rest of XPB-C may force the C-terminal extension to avoid the 'elbow'-like shape, which is its natural conformation in full-length XPB, in the crystal packing of XPB-C without helicase domain 1 (Fig. 4*a*). When XPB-C and the N-terminal half of AfXPB (PDB entry 2fz4) are separately superimposed with helicase domain 2 and helicase domain 1 of the RIG-I–dsRNA complex (PDB entry 2ykg; Luo *et al.*, 2011; Kowalinski *et al.*, 2011; Figs. 4*a* and 5), the position of the dsRNA is amazingly similar to that of the dsDNA/RNA in the model of the XPB–DNA complex, as shown in Fig. 3(*e*). Therefore, structural comparisons of XPB-C with Mss116p and RIG-I indicate almost identical XPB–DNA interactions. While the XPB–DNA complex derived from the Mss116p–dsDNA/RNA complex has issues with the C-terminal tail of XPB (Fig. 3*e*), replacing the C-terminal extension of XPB-C by the 'bridge' of RIG-I not only solves this problem but also predicts the position of the C-terminal tail of XPB (Fig. 5). Furthermore, an XPF–DNA complex (PDB entry 2bgw; Newman *et al.*, 2005) can be nicely docked onto this XPB–DNA complex to construct a structural model of the XPF–XPB–DNA complex for the 5′-incision reaction during nucleotide-excision repair (Fig. 5*b*). After XPB helicase (and XPD helicase) presumably unwinds the dsDNA around the lesion, XPF (and ERCC-1; not shown in the model) endonuclease binds to the ds–ssDNA junction to mediate the 5′ incision. Although the position of XPF in the complex is decided by the overlay of the dsDNA on the dsRNA, XPF is well positioned to interact with the C-terminal tail (at the position of the gray ribbons in Fig. 5) of XPB, consistent with biochemical results (Oksenych *et al.*, 2009).

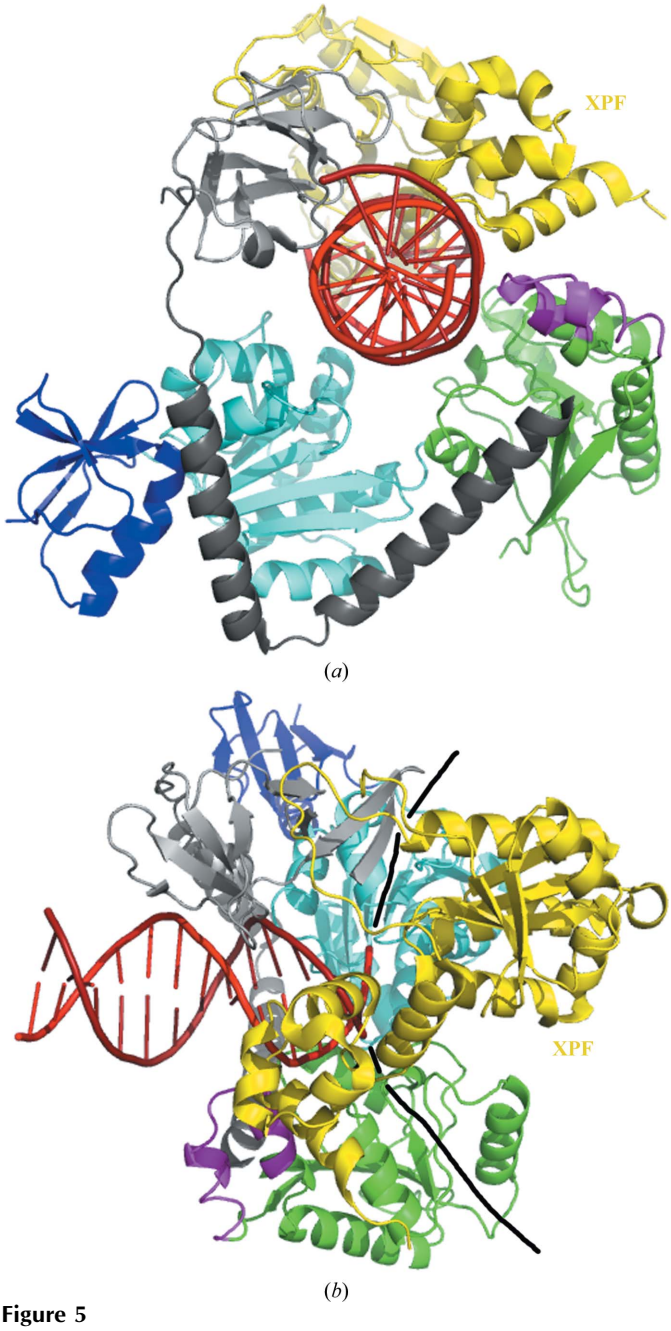


Figure 5
Structural model of the XPF–XPB–DNA complex. The side view (*a*) and the top view (*b*) of the XPF–XPB–DNA complex are shown in ribbon representation. The structural model was constructed by superimposing XPB-C and the N-terminal half of AfXPB (PDB entry 2fz4) with helicase domains 1 and 2 of RIG-I in the RIG-I–dsRNA complex (PDB entry 2ykg), respectively. The CTE of XPB-C is replaced by the 'bridge' of RIG-I. XPF and dsDNA were docked by superimposing the dsDNA of the archaeal XPF–DNA complex (PDB entry 2bgw) with the dsRNA of the RIG-I–dsRNA complex. In (*b*), the black lines extending from the DNA strands indicate possible paths for ssDNA unwound by XPB during DNA repair.

3.4. Effects of the XP11BE disease mutation

The current evidence supports the direct involvement of the C-terminus of XPB in DNA repair, but indicates that it is dispensable for transcription since phosphorylation of Ser751 prevents 5′ incision by the ERCC1-XPF endonuclease during

DNA repair but has no effect on transcription (Coin *et al.*, 2004). In addition, a C-terminal truncation of the yeast homolog Rad25 led to deficient DNA repair while retaining normal transcriptional activity (Guzder *et al.*, 1994). Similar results were also obtained with XPB$^{\Delta 43}$ mice, which produce a truncated XPB without the last 43 residues (Andressoo *et al.*, 2009). XPB$^{\Delta 43}$ mice failed to show a detectable CS-like developmental phenotype or accelerated segmental ageing features, but had a severe XP phenotype. These results indicate that the sequence change caused by the XP11BE mutation causes more damage than that caused by the loss of the last 41 residues. In order to find out why this mutation impairs transcription, we engineered the mutant XPBm(494–781), which contains a frameshift mutation that mimics the XP11BE mutation (Supplementary Fig. S1). To our surprise, the resulting GST-XPBm(494–781) was almost insoluble. Under the same conditions, the soluble level of GST-XPBm(494–781) is less than 10% that of GST-XPB-C (compare lanes T and S in Fig. 2*c* with lanes T and S in Fig. 2*a*), suggesting that the XP11BE mutation causes a reduction in the solubility of the XPB mutant produced in the XP11BE patient cells, which presumably leads to XPB protein degradation and reduces the intracellular level of XPB for the assembly of TFIIH in the XP11BE patient cells. To test this hypothesis, we used Western blotting to analyse the intracellular level of XPB and other TFIIH subunits, including XPD and p62, in lymphoblast cells obtained from the XP11BE patient and parents. It has been shown by electron microscopy (Schultz *et al.*, 2000) that XPB, XPD, p62, p52 and p44 form the ring-shaped core of the human TFIIH complex. Fig. 6 shows a representative result of Western blotting analysis of XPB, XPD and p62 (three subunits within the TFIIH complex) together with an internal control (vinculin). In comparison, the cellular levels of XPF and p53, a key DNA-repair protein, were also monitored. The mutant XPB from XP11BE patient cells could not be detected using a commercial antibody (XPB-sc) specific for the XPB C-terminus, but an antiserum against purified XPB-C was able to detect the reduced level of mutant XPB in the XP11BE patient (Fig. 6). This reduction was not caused by the antiserum, as it has the same reaction towards GST-XPB-C and GST-XPBm(494–781) in Western blotting (Supplementarry Fig. S2). The results showed that the cells from the XP11BE patient contained less than half of the XPB protein and the associated TFIIH subunits (XPD and p62) compared with those from the patient's healthy parents, suggesting that the assembly of the TFIIH complex is regulated by the intracellular level of XPB protein. In contrast, both XPF, the factor that directly interacts with XPB in DNA repair, and the tumor suppressor protein p53 remained at the same levels in all of the cells. The reduced level of intracellular XPB protein is not caused by transcription, since previous studies have shown no significant reduction in the level of XPB mRNA in XP11BE patient cells (Weeda *et al.*, 1990). Instead, the reduced levels of intracellular XPB protein and the TFIIH complex in the XP11BE patient are likely to be caused by reduced XPB solubility, as shown for the recombinant XPB mutant in *E. coli*.

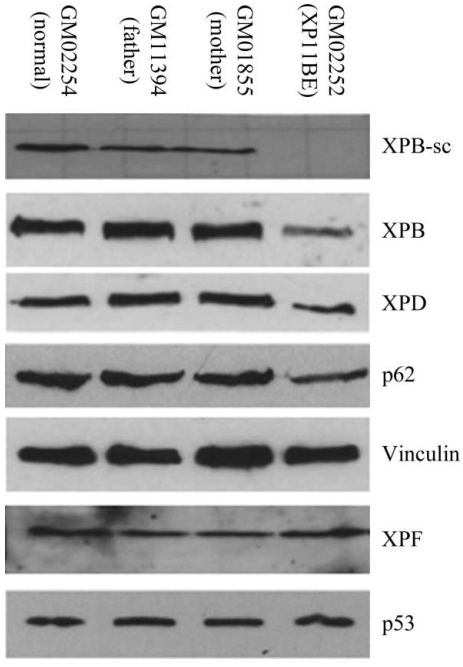


Figure 6
The XP11BE mutation reduces the intracellular levels of XPB and the TFIIH complex in the patient. Representative results of Western blotting. Lysates of lymphoblast cells from a patient with normal XPB (GM02254), the XP11BE patient (GM02252) and the mother (GM01855) and father (GM11394) of the XP11BE patient were analyzed by Western blotting for subunits of the TFIIH complex (XPB, XPD and p62), XPF, p53 and vinculin. XPB-sc indicates a Western blot with antibody specific for the C-terminus of XPB; XPB indicates a Western blot with antiserum against purified XPB-C protein.

In summary, we have determined the crystal structure of the C-terminal half of XPB at high resolution. The structural similarity of XPB-C to RIG-I suggests a possible model of the XPB–XPF–DNA complex for the 5′-incision reaction during nucleotide-excision repair. The XP11BE mutation not only divests the XPF-interaction motif, impairing DNA repair, but also reduces XPB solubility in the patient's cells, leading to a much lower level of intracellular XPB available for assembly of the TFIIH complex. The limited level of the TFIIH complex is in turn insufficient for genome-wide basal and activated transcription, so that the manifestation of CS was observed in the XP11BE patient. The results provide an explanation at the molecular level of the long-standing mystery of why the XP11BE mutation causes XP/CS complex manifestations.

This work was supported by a start-up fund to LF from UCR College of Natural and Agricultural Sciences and by the Hellman Fellowship Award to LF.

References

Adams, P. D *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Andressoo, J. O., Weeda, G., de Wit, J., Mitchell, J. R., Beems, R. B., van Steeg, H., van der Horst, G. T. & Hoeijmakers, J. H. (2009). *Mol. Cell. Biol.* **29**, 1276–1290.
Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* D**67**, 271–281.
Bootsma, D., Kraemer, K. H., Cleaver, J. E. & Hoeijmakers, J. H. J. (2002). *The Genetic Basis of Human Cancer*, 2nd ed., edited by B. Vogelstein & K. W. Kinzler, p. 821. New York: McGraw–Hill.

Brumback, R. A., Yoder, F. W., Andrews, A. D., Peck, G. L. & Robbins, J. H. (1978). *Arch. Neurol.* **35**, 337–345.

Bubeck, D., Reijns, M. A., Graham, S. C., Astell, K. R., Jones, E. Y. & Jackson, A. P. (2011). *Nucleic Acids Res.* **39**, 3652–3666.

Chapados, B. R., Chai, Q., Hosfield, D. J., Qiu, J., Shen, B. & Tainer, J. A. (2001). *J. Mol. Biol.* **307**, 541–556.

Cleaver, J. E., Thompson, L. H., Richardson, A. S. & States, J. C. (1999). *Hum. Mutat.* **14**, 9–22.

Coin, F., Auriol, J., Tapias, A., Clivio, P., Vermeulen, W. & Egly, J.-M. (2004). *EMBO J.* **23**, 4835–4846.

Compe, E. & Egly, J.-M. (2012). *Nature Rev. Mol. Cell Biol.* **13**, 343–354.

DeLano, W. L. (2002). *PyMOL*. http://www.pymol.org.

Douziech, M., Coin, F., Chipoulet, J.-M., Arai, Y., Ohkuma, Y., Egly, J.-M. & Coulombe, B. (2000). *Mol. Cell. Biol.* **20**, 8168–8177.

Dvir, A., Garrett, K. P., Chalut, C., Egly, J.-M., Conaway, J. W. & Conaway, R. C. (1996). *J. Biol. Chem.* **271**, 7245–7248.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.

Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.

Fan, L., Arvai, A. S., Cooper, P. K., Iwai, S., Hanaoka, F. & Tainer, J. A. (2006). *Mol. Cell*, **22**, 27–37.

Fan, L., Fuss, J. O., Cheng, Q. J., Arvai, A. S., Hammel, M., Roberts, V. A., Cooper, P. K. & Tainer, J. A. (2008). *Cell*, **133**, 789–800.

Gillet, L. C. & Schärer, O. D. (2006). *Chem. Rev.* **106**, 253–276.

Guzder, S. N., Qiu, H., Sommers, C. H., Sung, P., Prakash, L. & Prakash, S. (1994). *Nature (London)*, **367**, 91–94.

Hoeijmakers, J. H. (1994). *Eur. J. Cancer*, **30**A, 1912–1921.

Holm, L. & Rosenström, P. (2010). *Nucleic Acids Res.* **38**, W545–W549.

Holstege, F. C., van der Vliet, P. C. & Timmers, H. T. (1996). *EMBO J.* **15**, 1666–1677.

Huang, H.-R., Rowe, C. E., Mohr, S., Jiang, Y., Lambowitz, A. M. & Perlman, P. S. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 163–168.

Kim, T.-K., Ebright, R. H. & Reinberg, D. (2000). *Science*, **288**, 1418–1422.

Kowalinski, E., Lunardi, T., McCarthy, A. A., Louber, J., Brunel, J., Grigorov, B., Gerlier, D. & Cusack, S. (2011). *Cell*, **147**, 423–435.

Kraemer, K. H. (2003). *Fitzpatrick's Dermatology in General Medicine*, edited by I. M. Freedberg, A. Z. Eisen, K. Wolff, K. F. Austen, L. A. Goldsmith & S. I. Katz, pp. 1508–1521. New York: MacGraw–Hill.

Lehmann, A. R. (2003). *Biochimie*, **85**, 1101–1111.

Luo, D., Ding, S. C., Vela, A., Kohlway, A., Lindenbach, B. D. & Pyle, A. M. (2011). *Cell*, **147**, 409–422.

Mallam, A. L., Del Campo, M., Gilman, B., Sidote, D. J. & Lambowitz, A. M. (2012). *Nature (London)*, **490**, 121–125.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.

Mohr, G., Del Campo, M., Turner, K. G., Gilman, B., Wolf, R. Z. & Lambowitz, A. M. (2011). *J. Mol. Biol.* **413**, 952–972.

Moreland, R. J., Tirode, F., Yan, Q., Conaway, J. W., Egly, J.-M. & Conaway, R. C. (1999). *J. Biol. Chem.* **274**, 22127–22130.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.

Newman, M., Murray-Rust, J., Lally, J., Rudolf, J., Fadden, A., Knowles, P. P., White, M. F. & McDonald, N. Q. (2005). *EMBO J.* **24**, 895–905.

Noojin, R. O. (1965). *Arch. Dermatol.* **92**, 422–423.

O'Donovan, A., Davies, A. A., Moggs, J. G., West, S. C. & Wood, R. D. (1994). *Nature (London)*, **371**, 432–435.

Oh, K.-S., Khan, S. G., Jaspers, N. G. J., Raams, A., Ueda, T., Lehmann, A., Friedmann, P. S., Emmert, S., Gratchev, A., Lachlan, K., Lucassan, A., Baker, C. C. & Kraemer, K. H. (2006). *Hum. Mutat.* **27**, 1092–1103.

Oksenych, V., Bernardes de Jesus, B., Zhovmer, A., Egly, J.-M. & Coin, F. (2009). *EMBO J.* **28**, 2971–2980.

Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.

Rapin, I., Lindenbaum, Y., Dickson, D. W., Kraemer, K. H. & Robbins, J. H. (2000). *Neurology*, **55**, 1442–1449.

Riou, L., Zeng, L., Chevallier-Lagente, O., Stary, A., Nikaido, O., Taïeb, A., Weeda, G., Mezzina, M. & Sarasin, A. (1999). *Hum. Mol. Genet.* **8**, 1125–1133.

Robbins, J. H., Kraemer, K. H., Lutzner, M. A., Festoff, B. W. & Coon, H. G. (1974). *Ann. Intern. Med.* **80**, 221–248.

Schaeffer, L., Roy, R., Humbert, S., Moncollin, V., Vermeulen, W., Hoeijmakers, J. H., Chambon, P. & Egly, J.-M. (1993). *Science*, **260**, 58–63.

Schultz, P., Fribourg, S., Poterszman, A., Mallouh, V., Moras, D. & Egly, J.-M. (2000). *Cell*, **102**, 599–607.

Scott, R. J., Itin, P., Kleijer, W. J., Kolb, K., Arlett, C. & Muller, H. (1993). *J. Am. Acad. Dermatol.* **29**, 883–889.

Shivji, M. K. K., Podust, V. N., Hubscher, U. & Wood, R. D. (1995). *Biochemistry*, **34**, 5011–5017.

Sijbers, A. M., de Laat, W. L., Ariza, R. R., Biggerstaff, M., Wei, Y.-F., Moggs, J. G., Carter, K. C., Shell, B. K., Evans, E., de Jong, M. C., Rademakers, S., de Rooij, J., Jaspers, N. G. J., Hoeijmakers, J. H. J. & Wood, R. D. (1996). *Cell*, **86**, 811–822.

Vermeulen, W., Scott, R. J., Rodgers, S., Müller, H. J., Cole, J., Arlett, C. F., Kleijer, W. J., Bootsma, D., Hoeijmakers, J. H. & Weeda, G. (1994). *Am. J. Hum. Genet.* **54**, 191–200.

Weeda, G., Eveno, E., Donker, I., Vermeulen, W., Chevallier-Lagente, O., Taïeb, A., Stary, A., Hoeijmakers, J. H., Mezzina, M. & Sarasin, A. (1997). *Am. J. Hum. Genet.* **60**, 320–329.

Weeda, G., van Ham, R. C., Vermeulen, W., Bootsma, D., van der Eb, A. J. & Hoeijmakers, J. H. (1990). *Cell*, **62**, 777–791.

Wood, R. D. (1997). *J. Biol. Chem.* **272**, 23465–23468.

Yoneyama, M., Kikuchi, M., Matsumoto, K., Imaizumi, T., Miyagishi, M., Taira, K., Foy, E., Loo, Y.-M., Gale, M., Akira, S., Yonehara, S., Kato, A. & Fujita, T. (2005). *J. Immunol.* **175**, 2851–2858.

Yoneyama, M., Kikuchi, M., Natsukawa, T., Shinobu, N., Imaizumi, T., Miyagishi, M., Taira, K., Akira, S. & Fujita, T. (2004). *Nature Immunol.* **5**, 730–737.