HHS Public Access

Author manuscript

Nature. Author manuscript; available in PMC 2023 August 11.

Published in final edited form as:

Nature. 2023 May ; 617(7959): 170–175. doi:10.1038/s41586-023-05959-z.

Lesion recognition by XPC, TFIIH and XPA in DNA Excision Repair

Jinseok Kim[1], **Chia-Lung Li**[1], **Xuemin Chen**[1], **Yanxiang Cui**[2], **Filip M. Golebiowski**[1,5], **Huaibin Wang**[2], **Fumio Hanaoka**[3], **Kaoru Sugasawa, PhD**[4], **Wei Yang, PhD**[1]

[1]Laboratory of Molecular Biology, NIDDK, National Institutes of Health, Bethesda, MD 20892

[2]Laboratory of Cell and Molecular Biology, NIDDK, National Institutes of Health, Bethesda, MD 20892

[3]National Institute of Genetics, Research Organization of Information and Systems, Mishima, Shizuoka 411-8540, Japan

[4]Biosignal Research Center, and Graduate School of Science, Kobe University, Kobe, Hyogo 657-8501, Japan.

Abstract

Nucleotide excision repair (NER) removes DNA lesions caused by ultraviolet light (UV), cisplatin-like compounds and bulky adducts[1,2]. After initial recognition by XPC in global genome repair (GGR) or a stalled RNA polymerase in transcription-coupled repair (TCR), damaged DNA is transferred to the seven-subunit TFIIH core complex (Core7) for verification and dual incisions by the XPF and XPG nucleases 3. Structures capturing lesion recognition by the yeast XPC-homolog Rad4 and TFIIH in transcription initiation or DNA repair have been separately reported[4–8]. How two different lesion-recognition pathways converge and how the XPB and XPD helicases of Core7 move lesion DNA for verification are unclear. We report here structures revealing DNA-lesion recognition by human XPC and DNA-lesion handoff from XPC to Core7 and XPA. XPA, which binds between XPB and XPD, kinks the DNA duplex and shifts XPC and the DNA lesion by nearly a helical turn relative to Core7. The DNA lesion is thus positioned outside of Core7, as would occur with RNA polymerase. XPB and XPD, which track the lesion-containing strand but translocate DNA in opposite directions, push and pull the lesion-containing strand into XPD for verification.

Keywords

Nucleotide excision repair (NER); Xeroderma pigmentosum; Cockayne syndrome; ds-ss junction

Corresponding Authors: Wei Yang, PhD weiy@niddk.nih.gov, Kaoru Sugasawa, PhD ksugasawa@garnet.kobe-u.ac.jp.

[5]Current address: Roche Polska, 52J4+Q6 Warsaw, Poland

Author contribution

J.K. carried out biochemical and structural studies; C.L.L. carried out dual incision assays; F.M.G. developed the GraFix protocol; H.W. and Y.C. helped with cryoEM grid preparation and data acquisition, Y.C. and X.C. helped with cryoEM data processing and map improvement; K.S. and W.Y. conceived the research project; W.Y. supervised experimental design and data interpretation; K.S. and F.H. helped with data interpretation; all authors were involved in writing the paper and adhere to the “Inclusion & Ethics” regulation.

Competing interests

The authors declare no competing interest.

Introduction

NER is essential for maintaining genome integrity. Defects in NER are implicated in human autosomal recessive disorders: Xeroderma Pigmentosum (XP), characterized by extreme sensitivity to UV and skin lesions, Cockayne syndrome (CS) and Trichothiodystrophy (TTD), marked by delayed and impaired development, mental retardation and premature death[2,9,10]. The XPC complex is trimeric and composed of XPC (Xeroderma Pigmentosum complementation group C), RAD23 and CETN2 (Centrin 2) subunit[11,12]. The XPC-RAD23 heterodimer binds to a variety of lesion sites with nanomolar affinity[13]. Crystal structures of the yeast XPC homolog (Rad4) and Rad23 complexed with damaged DNA show that Rad4 generates a 2-nt bubble by flipping out the lesion, whether it is a mismatched base pair, abasic site (AP) or UV-induced base dimer, and binds the unpaired non-lesion strand and DNA duplex 3¢ of the lesion[4]. CETN2, which enhances lesion recognition and GGR[14], is absent in all published Rad4-DNA structures[4,15,16].

TFIIH (transcription factor II H) contains two helicases, XPB and XPD. Only the ATP-dependent dsDNA translocase activity of XPB is necessary for transcription initiation[17]. In the transcription pre-initiation complex (PIC), XPD serves as a structural partner that bridges RNA polymerase and TFIIH, while its DNA binding and helicase activity are blocked by the CAK kinase module of TFIIH (composed of CDK7, MAT1 and Cyclin H)[6–8,18,19]. By contrast, during NER the inhibitory CAK module dissociates upon XPA binding[20], and the XPD helicase activity is essential for scanning ssDNA in the 5′ to 3′ direction for lesion verification[21–23]. In the structure of Core7 and the NER-specific protein XPA complexed with a Y-shaped ds-ss DNA junction without any lesion, XPB and XPD bind ds- and ss-DNA, respectively, and XPA is wedged between the two helicases at the ds-ss junction[5]. The latest 9 Å resolution cryoEM structure of yeast Core7 and Rad4 complexed with a lesion DNA (Extended Date Fig. 1) suggests that the DNA lesion is placed between XPB and XPD, and thereby 5¢ to XPD[16]. If so, the 5¢ to 3¢ translocation activity of XPD would result in XPD moving away from the lesion and being unable to verify lesions.

We have determined cryoEM structures of human XPC bound to DNA containing the bulky lesion mimetic Cy5 (a fluorescent dye) or two consecutive abasic sites (AP) in the presence of Core7, and four additional complexes revealing XPA recruitment and displacement of XPC and the DNA lesion from Core7 (Extended Date Fig. 2–3). These structures reveal not only the intricacy of step-by-step recruitment of NER components, but also the unexpected yet logical initial placement of lesions outside of Core7, so XPB and XPD with opposite translocation polarity (3¢ to 5¢ and 5¢ to 3¢) together can scan DNA lesion in GGR and TCR alike.

DNA lesion recognition complexes

We expressed and purified human trimeric XPC, XPA and Core7 of TFIIH (seven subunits, XPB, XPD, p62, p52, p44, p34 and p8, also known as TTDA) separately (Fig.1a–b). A Cy5 paired with two thymines and flanked by 29 and 25 bp upstream and downstream, respectively, was used as a bulky lesion mimic. Formation of protein-DNA complexes was

confirmed by electrophoretic mobility shift assays (EMSA). Human trimeric XPC binds the Cy5-DNA with a Kd of <5 nM (Fig. 1c), comparable to published values for human XPC Kim et al., 2022 and yeast Rad4[13,14,24]. Although Core7 alone or with XPA binds the same DNA very weakly, Core7 and XPA together formed a distinct super-shifted band with XPC-DNA[23] (Fig. 1c). We used GraFix[25] to stabilize the complexes of XPC, XPA and Core7 with the Cy5-DNA and by cryoEM identified C7CD (for Core7, XPC and DNA), C7CAD (with additional XPA), and C7AD (with XPC displaced but not fully dissociated) complexes (Fig. 1D–F). These complexes co-existed in the absence of ATP and repair factors RPA, XPF and XPG. Their structures have been resolved at resolutions of 3.4 to 4.1 Å (Extended Data Fig. 2 and Table 1) and reveal the process of lesion recognition by XPC and subsequent handoff to TFIIH and XPA.

The previously unknown structure of trimeric human XPC was determined at 4.2 Å in the C7CD complex. Due to the mobility, peripheral domains and DNA ends are at lower resolutions (Extended Data Fig. 2d). In addition to a 3-nt DNA bubble at the lesion site, the C7CD structure reveals how XPC (indicating the XPC subunit hereafter unless specified as trimeric XPC) orients Core7 to bind the lesion DNA (Fig. 1d). The U-shaped Core7 is related to the apo-TFIIH structure (PDB: 6NMI)[19] and human and yeast PIC (PDB: 7NVW and 7ML0)[6,8]. Only XPB binds the DNA upstream (5¢) of the lesion. XPD is over 20 Å away from the DNA on the downstream side of Cy5. If present, the CAK module would not interfere with the C7CD formation.

In the C7CAD complex with XPA and without CAK[20], Core7 is transformed from the U shape to a closed σ shape, and XPD contacts both XPB and DNA (Fig. 1e). XPC maintains its binding of the Cy5 and downstream DNA duplex (Fig. 1e). The presence of XPA moves Core7 one helical turn away from XPC and shifts the Cy5 from above the 5 crevice between XPB and XPD in C7CD to outside of Core7 in C7CAD. The DNA duplex between XPA and XPD is kinked and locally unwound 7 bp upstream of the Cy5 and primed to form a ds-ss junction.

In the C7AD structure, the trimeric XPC is largely detached from Core7 except for the C- terminal 52 residues of XPC (Fig. 1f). The Cy5 and surrounding DNA are disordered due to their association with the detached and mobile XPC and incomplete transfer to XPD. Core7, XPA and the upstream DNA in C7AD are configured same as in C7CAD and superimposable with the structure of XPA and Core7 bound to a splayed non-lesion DNA (PDB: 6RO4)[5] except for only 2 nt of ssDNA visibly bound to XPD. We cannot rule out that some C7AD complexes may have resulted from Core7 and XPA binding to dsDNA ends instead of XPC-associated lesions.

DNA with two consecutive abasic sites (AP) replacing Cy5 as a lesion substrate (AP-DNA) also formed three complexes as observed with Cy5-DNA. These structures were determined at 3.6 to 7.1 Å resolution (Extended Data Fig. 3 and Table 1) and share general features with the Cy5 complexes. Descriptions below are based on the Cy5 complexes with unique features of the AP-DNA complexes reported.

Flexible XPC and lesion-DNA association

The human XPC-DNA complexes without Core7 appeared very flexible on cryoEM grids and yielded no meaningful structures. As shown in cellular studies[26], the trimeric XPC structure is stabilized by TFIIH on lesion DNAs and becomes well defined in C7CD. Similar to Rad4, XPC has four globular domains: one TGD (transglutaminase-homology domain) and three BHDs (β-hairpin domain, 1–3) (Fig. 2a). AlphaFold[27] provided a more accurate human XPC model than the Rad4 crystal structure, due to less than 27% sequence identity and multiple deletions and insertions. With nearly 200 amino acids (aa) more than Rad4, human XPC (940 residues) contains four short deletions and a single large insertion of 180 residues in the TGD (Extended Data Fig. 4). This large insertion, composed of >60% Gly, Ser, Lys, Arg and Pro, is unstructured, but its positive charge and proximity to the DNA suggest a role in DNA binding. This region is also known to interact with a non-acetylated histone H3 tail and promote GGR[28]. As with Rad4, only a four-helix domain (~60 residues) of RAD23 is traceable and forms a hydrophobic interface with TGD (Fig. 1a, 2a). The remaining RAD23 is involved in ubiquitin binding and XPC recruitment to UV lesions[29].

Many DNA-binding features are conserved between XPC and Rad4. Among the three subunits, only XPC directly contacts the DNA, while RAD23 and CETN2 stabilize the XPC structure and enhance XPC-DNA association (Fig. 2a). XPC binds DNA chiefly along one side of the duplex with TGD and BHD1 covering 7 bp (+5 to +11) downstream of the Cy5. The β hairpins of BHD2 (β22-β23, P703-F733) and BHD3 (β32-β33, I794-D809) approach the non-lesion DNA strand from the minor and major groove, respectively (Fig. 2b). Together they create a DNA bubble by displacing the lesion and flipping out bases on the non-lesion strand, two by Rad4 and three by XPC. T26 and T27 opposite the Cy5 and the adjacent A28 are inserted into XPC (Fig. 2b). T26 appears sandwiched between Y656 and H685 (Extended Data Fig. 5a), which are unique in XPC. T27 and A28, which are situated similarly to the bases flipped out by Rad4, are surrounded by conserved F762, F797, F799, F756, I794 and P806 (Extended Data Fig. 4).

Interestingly, the AP_C7CD complex adopts two conformations (Extended Data Fig. 3). One has three bases flipped out as in Cy5_C7CD, and the other has two bases opposite the AP site (T26 and T27) flipped into the conserved pockets that accommodate T27 and A28 in the Cy5 structure (Fig. 2c). Between yeast and human, each structural domain of XPC/Rad4 and RAD23 can be individually superimposed, but their relative orientations differ significantly (Extended Data Fig. 6a–f). Likewise, between flipping out 2 and 3 nt in the non-lesion strand, RAD23 and domains in XPC re-orient and translocate along DNA by one base pair (video 1). The TGD distal from the DNA interface moves 25 Å. The remarkable flexibility within trimeric XPC and between XPC and DNA results in reduced structure resolution but likely underlies the high affinity of XPC for a variety of DNA lesions, large and small.

XPC also exhibits unique features. In contrast to exposed and disordered lesions when complexed with Rad4, the Cy5 appears ordered. Although displaced from the DNA and the BHD domains, the aromatic moiety of Cy5 appears surrounded by hydrophobic residues (I182, F186 and Y189) on a long α-helix (aa 167–224) N-terminal to the TGD domain, which is named LHN (Fig. 2b). LHN extends to the XPD arm of Core7 in C7CD. Without

Core7, LHN of Rad4 is shortened and broken in two, both of which bind DNA backbones with positively charged sidechains[4]. Next to Cy5, R192 and R196 contact the non-lesion strand phosphates (Extended Data Fig. 5b). An R192 to Gly mutation in germline cells is found in young onset prostate cancer patients[30]. Notably, neither the LHN hydrophobic residues nor R192 are conserved in Rad4 (Extended Data Fig. 4). Another long α helix in XPC (aa 815–866) at the C-terminus of BHD3, thus named LHC, extends toward Core7 and crosses the DNA before contacting XPB. LHC is stabilized by the C-terminal EF-hand domain of CETN2, which enhances XPC-DNA association[14] (Fig. 2a). CETN2 alone can bind LHC (aa 837–853)[31,32] (Extended Data Fig. 6g–h). With LHC anchored onto Core7 in C7CD, the preceding BHD3 interacts extensively with the 8-bp DNA upstream of the lesion (Fig. 2a–b). In the absence of Core7, LHC of Rad4 is disordered, and the DNA is severely bent (Fig. 2d) away from BHD3 in the crystal structure[4].

Recruitment of Core7 and XPA by XPC

In C7CD, the lesion DNA is sandwiched between XPC and Core7 (Fig. 3a). Abutting BHD3, XPB binds 11 bp DNA upstream of the Cy5 (−5 to −15 bp), burying 1158 Å^2 at the interface. The DNA is bent toward XPB with a widened minor groove. Similar DNA distortion by XPB is also observed in the transcriptional PIC 7 (Extended Data Fig. 7a). Two long helices LHN and LHC of XPC, 45 Å apart, extend from XPC to Core7 like bridges across the DNA. The N-terminus of LHN (aa 175–186) contacts p62 (aa 175–186) on the XPD arm, forming a small interface (250 Å2). Although disordered in C7CD (due to flexible inter-domain linkers), the XPC and p62 interface is likely extended by 1000 Å^2 between the acidic stretch of XPC (aa 109–156) and the PH (Pleckstrin Homology) domain of p62 (aa 1–108) as reported previously[33,34]. The C-terminus of LHC (aa 853–866) is anchored by forming a pair of anti-parallel helices with XPB (aa 706–718) (Fig. 3a). Two following short helices in XPC (L and M, aa 891–913) form additional interactions with XPB, p52 and p8. Mutation of A896 to Val in the center of this interface is linked to predisposition to prostate cancers[30]. Recruitment of Core7 is oriented by LHN and LHC, and the attachment appears to support the ~30° movement of XPC between flipping out 2 and 3 nt (Video 1).

When XPA is present in C7CAD, XPB and Core7 are shifted upstream one helical turn away from Cy5, which remains bound to XPC (Video 2). The vacated DNA is occupied by XPA and XPD (Fig. 3b–d). The N-terminal 100 residues of XPA, which interact with RPA[35,36], are disordered (Fig. 1a). The central αβ domain of XPA (aa 137–196) approaches the adjacent minor and major grooves opposite Core7 (Fig. 3b). W175 on the XPA β-hairpin is intercalated between adjacent base pairs −7 and −8 from the Cy5 (Fig. 3d, Extended Data Fig. 5c), causing DNA unwinding and kinking 50° toward Core7. HD2 (helicase domain 2) of XPD binds the lesion strand around the kink (Fig. 3b).

XPA bears an uncanny similarity to the BHD3 of XPC in its tertiary structure and DNA-binding mode. The αβ domain of XPA and BHD3 (XPC) are separated by 35 Å (equal to a full helical turn) in C7CAD (Fig. 3b). But due to the severe DNA unwinding at the lesion site, the β-hairpin of XPA contacts the minor groove, while the β-hairpin of BHD3 contacts the adjacent major groove. The XPA structure is superimposable with its apo form[37] and in complexes with DNA[5,38,39]. However, unlike previously reported XPA-DNA complexes,

XPA binds the DNA duplex in C7CAD and not a ds-ss junction or DNA end (Fig. 1e and 3b).

A long helix following the αβ domain in XPA (aa 196–233), named LHA, replaces LHC (XPC) and is inserted between XPC and DNA in C7CAD (Fig. 1e, 3b). The extended LHA first binds the DNA across the minor groove and then reaches the XPB arm of Core7. The linker (aa 234–243) and C-terminal β domain of XPA (aa 244–273) consecutively interact with XPB, p8 and p52 (Fig. 3c, 3e), as suggested in previous analyses[5]. Anchoring of XPA in the XPB arm appears important as mutations (including H244R) or truncation of the last ~50 residues of XPA are implicated in UV sensitivity and neurological impairment[40–43]. LHA and LHC cross one another at an ~40° angle, and CETN2 appears to stabilize the small XPA-XPC interface at the crossover (Fig. 3b), which corroborates previous biochemical observations[44].

In the presence of XPA, the C-terminus of XPC is ordered and traceable to the penultimate residue K393 (Fig. 3f). The last 50 residues of XPC forming helices L, M and N are anchored in the HD2 of XPB, p52, p8 and p34, approximately 35 Å away from the XPA attachment site, thus enabling the transition of C7CD to C7CAD (Video 2). A truncated XPC with a K928Q mutation is implicated in cancer susceptibility[45]. K939 caps the C-terminus of the preceding helix L (aa 924–933) (Extended Data Fig. 5d), and the polymorphic variation K939Q, which is implicated in bladder cancers[46], may perturb the XPC structure and weaken the XPC and Core7 association. When Cy5 is replaced by two abasic sites, the structure of Core7 in C7CAD changes very little, while the AP-DNA is overall more bent (Fig. 3g). Locally AP-DNA bound by XPB is less bent than Cy5-DNA. Concomitantly, both LHA and LHC move slightly away from DNA (Video 3).

Modular and flexible TFIIH

Core7 undergoes significant conformational changes from C7CD to C7CAD, with the XPB and XPD arms moving closer and crossing each other (Video 4–5). XPB is attached to p34 via p52 and p8, and XPD to p34 via p44 and p62 (Extended Data Fig. 8a–c). While the interfaces of p34 with p52 and p44 remain unchanged, the modular structures and domain flexibility in p52 and p44 underlie the movements of the XPB and XPD arms (Video 6–7). In C7CD, the XPD arm is secured by the extended p62, which wraps around p52-p34-p44 and also around the unattached XPD-p44. The XPB arm is fastened by the four-domain p52 (see below) and by XPC and XPA at the interface of XPB, p52, p8 and p34 (Fig. 3b–f). In C7CAD, the two arms of Core7 are united by XPD binding to XPB and DNA and by p34 and p44 (aa 8–54) interacting with the first three domains of XPB (Fig. 1a, Extended Data Fig. 8b).

The five non-enzymatic subunits in TFIIH are modular and repetitive. p34 and p44 both contain a central vWA domain, and each has additional zinc finger domains (Fig. 1a). The p34-vWA domain forms a stable interface with the C-terminal RING finger domain of p44 (RF), while the C-terminal double zinc finger domain of p34 (DZF) and the zinc ribbon domain of p44 (ZR) form a highly flexible joint in the XPD arm (Extended Data Fig. 8c–d). DZF of p34, which is all β strands in C7CD, is altered to contain an α helix in C7CAD.

p52 contains four domains. Two at the N-terminus resemble one another and form a stable interface with the vWA domain of p34. Two at the C-terminus fold like DRD (Damage Recognition Domain), as do their interacting partners, the N-terminal DRD of XPB[47] and p8 (Extended Data Fig. 9 and 8e–f). Despite the name, none of DRD domains in Core7 interacts with DNA. We note that the heterodimer formed by p52 and p8 (Fig. 3f, Extended Data Fig. 9) is superimposable with the p8 homodimer[48] without detectable sequence homology. The p8 dimer may partially rescue the lethal defects of deleting domains 3 and 4 of p52 in Drosophila[49]. The small p8 functions like glue to keep the dynamic Core7 together and also tie down XPA[42] (Fig. 3b–c, f).

Coordination of XPB and XPD motors in NER

The C7AD structure is nearly identical to that of Core7 and XPA complexed with a splayed non-lesion DNA[5]. The DNA strands are separated by XPA, where W175 stacks with the last base pair (Fig. 4a–b), and only two nucleotides of the lesion strand enter the DNA-binding groove of XPD. The bulk of XPC and the associated DNA lesion are disordered, and only the C-terminal helices of XPC remain bound to Core7. The absence of ATP and other NER factors probably prevents XPC from releasing the lesion DNA. Based on the structures of C7AD and Core7 and XPA with the non-lesion DNA (PDB: 6RO4)[5], after XPC's dissociation XPD helicase will bind and scan the lesion strand. In addition to the Fe_4S_4 cluster proposed to detect DNA lesions[50–52], the ssDNA exit in XPD adjacent to XPA is narrow and lined with aromatic side chains (Fig. 4b) that may allow normal DNA through but stall on DNA lesions.

Among superfamily 2 (SF2) helicases, XPB, the yeast homolog of CSB (Cockayne syndrome group B) (PDB: 5VVR), and the chromatin remodeler Snf2 (PDB: 6IY2)[53,54] translocate dsDNA. These helicases and associated DNA duplexes are superimposable. When the viral SF2 helicase NS3, which binds and translocates ssRNA (PDB: 3O8C)[55], is included in the superposition, the translocating (or tracking) strand associated with all three dsDNA translocases overlays with the ssRNA (Extended Data Fig. 6b). While translocating DNA duplex, XPB tracks only one strand with its HD1 and HD2 (Helicase Domains 1 and 2), which is the lesion strand in NER (Fig. 4a) and DNA template strand in transcription[6]. If DNA were immobile, XPB would move along the lesion strand in the 3′ to 5′ direction and thus away from the Cy5 lesion. When constrained in the repair or transcription machinery, XPB would be stationary and translocate the DNA duplex toward XPA and XPD in NER (Fig. 4a) and RNA polymerase in PIC.

As shown in previous studies, XPD is a stronger motor than XPB and dictates the direction of Core7 movement 5¢ to 3¢ along the lesion strand when both motors are active[22,23]. We hypothesize that XPD translocates along the lesion strand, pulling it into Core7 until it is stalled by a DNA lesion. When XPD is stuck, XPB becomes the only working motor and pushes the DNA upstream of the lesion toward XPA and XPD (Fig. 5). Armed with the duplex-separation β-hairpin and W175, XPA separates the DNA duplex pushed toward it by XPB and enlarges the DNA bubble. When the DNA bubble reaches 24–34 nt in size[56], dual incision by XPF and XPG takes place.

To test this hypothesis, we developed an in vitro dual incision assay, replacing circular plasmid DNA substrate[22,57,58] with a linear 94 bp DNA containing either a Cy5 or two consecutive AP sites. DNA cleavage occurred after addition of ATP, XPF, XPG and RPA (Fig. 4c). As predicted for the XPB-driven bubble formation, a 62-bp duplex upstream to the lesion leads to efficient dual incision at ~20 nt 5¢ and <10 nt 3¢ to the lesion. Shortening the upstream duplex to 52 bp reduced the dual incision by six-fold (Extended Data Fig. 10). Removal of Cy5 is more efficient than AP, confirming that XPD discriminates between different lesions. As reported previously[59], XPG can cleave 3¢ to the lesion in the absence of XPF, but the 5¢ incision by XPF requires the presence of XPG (Fig. 4c).

Conclusion

Damaged DNA has a propensity to bend and deform and is recognized by XPC via a 2 or 3 nt bubble in which normal bases in the non-lesion strand are specifically bound by XPC. With LHN and LHC, XPC recruits and, importantly, orients Core7 so that XPD contacts the lesion strand. XPA plays a key role in shifting Core7 relative to XPC-lesion DNA, so the lesion falls downstream and 3′ to XPD (Fig. 5)[7,8]. Although it is unknown how TFIIH is recruited to a stalled transcription site, TFIIH likely interacts with RNA polymerase as in PIC with the lesion between them. After XPA dissociates the CAK module[20], Core7 would bind the lesion DNA as if recruited by XPC. Subsequent NER reactions would occur as outlined for GGR and TCR alike (Fig. 5).

Methods

Protein expression and purification

The recombinant human Core7 (XPB, XPD-Flag and p8, p62 and p34, p52 and His-p44)[23], Flag-XPC[22] were expressed in Sf9, and RAD23B-His[60], Trx-His-centrin2[14], RPA[61] and XPA (Trx-His-XPA) in E. coli cells, and each was purified as previously described. To reconstitute trimeric XPC, 0.4 L Sf9 suspension culture expressing Flag-XPC was pelleted, mixed with 20 mg Rad23-His and 10 mg Centrin2 (separately purified) in 25 ml of the lysis buffer (50 mM Tris-HCl pH 7.5, 300 mM KCl, 20 mM Imidazole, 5% Glycerol, 10 mM β-mercaptoethanol (bME), and 0.01% Triton X-100) with 0.25 mM PMSF, 10 ug/ml Pepstatin A, Complete Protease Inhibitor Cocktail tablet [EDTA-free, Roche]) and lysed by sonication. The lysate was cleared by centrifugation (35,000 × g) for 30 min at 4°C and loaded onto a HisTrap column (5 ml, GE Healthcare Bioscience). After washing with 1 M KCl in the lysis buffer, the XPC trimer was eluted with 120 mM Imidazole in the lysis buffer. The eluent was mixed with equal volume of M2-wash buffer (50 mM Tris-HCl (pH 7.5), 300 mM KCl, 5% glycerol, 0.007% Triton X-100, 0.25 mM PMSF), and incubated with 8 ml anti-FLAG M2 agarose beads (Sigma) at 4°C for 1 hour. After extensive wash, trimeric XPC was eluted by 0.3 mg/ml 1X-FLAG peptide (Sigma) in the wash buffer with 5 mM bME and further purified through a Heparin column (5ml, GE Healthcare Bioscience) equilibrated with the Heparin-binding buffer (50 mM Tris-HCl pH 7.5, 300 mM KCl, 5% Glycerol, 10 mM bME, 0.007% Triton X-100, 0.25 mM PMSF). The bound XPC protein was eluted with a 75 ml-linear gradient of 0.3 to 1 M KCl. The XPC protein was concentrated and passed through a Superose 6 column (10/300, GE Healthcare

Bioscience) in 25mM Tris-HCl (pH7.4), 300 mM KCl, 5 mM DTT, 10% Glycerol, 0.007% Triton X-100. The eluted peak fractions were pooled, concentrated and stored at −80°C. The C-terminal MBP tagged XPG (XPG-MBP), XPF and the N-terminal His_6 and MBP tandemly tagged ERCC1 (His-MBP-ERCC1) were cloned into the pLEXm vector[62]. XPG and XPF/ERCC1 were separately expressed in HEK293 cells and purified using an amylose affinity column. MPB tags were removed by PreScission Protease treatment. XPG was further purified on Mono Q (10/100 GL, GE Healthcare Bioscience), and the eluant in 50 mM Tris-HCl pH 8.0, 400 mM NaCl, 5 mM DTT, 10% glycerol was concentrated, flash frozen, and stored at −80°C. After removing the MBP tag, XPF/ERCC1 was concentrated and purified on a Superdex 200 column (16/60, GE Healthcare Bioscience) with buffer containing 20 mM Tris-HCl pH8.0, 300 mM NaCl, 10% glycerol and 5 mM DTT. To further purify XPF/ERCC1 complex, the peak fraction from Superdex 200 column was loaded onto a 5ml Heparin column equilibrated with buffer (50 mM Tris- HCl pH8.0, 10% glycerol, 5 mM DTT). The XPF/ERCC1 complex was eluted with a 120 ml linear gradient of 0.05 to 1 M NaCl and the peak fractionations were pooled, concentrated, flash frozen, and stored at −80°C.

DNA substrates

All oligonucleotides were synthesized by Integrated DNA Technologies (IDT) and purified in-house using denaturing PAGE (Supplementary Information Table 1). For structural studies, Cy5 and 2AP-containing oligos (Top56_iCy5 and Top56_2AP) were each annealed with complementary Bot56 at 1:1 molar ratio at 20 μM concentration by heating to 95°C and slow cooling to 4°C. Resulting DNA duplexes were purified from excess ssDNA using native PAGE. For dual incision assays, the Cy5 strand was prepared by ligating Top_F1_47, Top_F2_iCy5, and Top_F3_10_5P oligos after annealing with Bot_F1_74 (Supplementary Table 1). The top strand containing an internal Cy5 (94 nt) was gel purified and annealed with the bottom strand mm94_Bot_62/30, which resulted in a 94 bp duplex with three mismatched base pairs 5' of the Cy5 to enhance dual incision activity. To make 2AP DNA, Top_F2_2AP was ^{32}P labeled at the 5′ end and ligated with Top_F1_53 and Top_F3_21_5P after annealing with Bot_F1_74 and gel purified to generate the 94 nt DNA oligo with two abasic sites (2AP) and the internal isotope label. The same bottom strand mm94_Bot_62/30 was annealed with the 2AP oligo.

CryoEM preparation and data collection

NER initiation complex was assembled by mixing 0.6 μM lesion-containing duplex DNA (56B2_iCy5 or 56B2_2AP, Supplementary Table 1) with Core7, XPC, and XPA at 1:1:1:1.5 molar ratio in 25 mM HEPES (pH 7.9), 100 mM KCl, 2 mM TCEP, 10% glycerol and incubation at 25°C for 10 min. These protein:DNA complexes were stabilized and purified using the GraFix method[25]. A glycerol (15–30%) and glutaraldehyde (0–0.04%) gradient was prepared with BioComp Gradient Master 108, and centrifuged at 165,000 g for 14 h at 4°C. The fractions containing the protein-DNA complexes were pooled and treated with 100 mM (final) Tris-HCl (pH 8.0) to quench the excess glutaraldehyde. The protein complex sample was dialyzed against 25 mM HEPES (pH7.9), 100 mM KCl, 2 mM $MgCl_2$, 2 mM TCEP, 0.5% glycerol in a Slide-A-Lyzer MINI (2 ml, 20 kDa MWCO) for 12 h and concentrated to 0.4 mg/ml for cryoEM grid preparation directly. A home-made 3-nm

carbon film was layered on Quantifoil (R1.2/1.3) holey carbon grids (EMS), which were glow-discharged for 30 seconds in an easy-glow system before usage. A 2 μl drop of each protein-DNA sample was applied to the grid in an FEI Vitrobot IV at 4°C with 100% humidity. The grid was blotted for 4 s with force 4 and plunge-frozen in liquid ethane cooled by liquid nitrogen. The Titan Krios G3 electron microscope (Thermofisher) equipped with a Gatan imaging filter (GIF) Quantum system and a post-GIF Gatan K3 Summit direct electron detector (Multi-Institute Cryo-EM Facility (MICEF)) of NIH operated at 300 kV with SerialEM (version 3.8.4)[63] was used to collect 6243 movies of Cy5_C7CAD and 3883 movies of 2AP_C7CAD at 105K nominal magnification in super- resolution mode (corresponding to calibrated pixel size 0.4165Å at the sample level). During data collection, each movie was recorded with an exposure time of 2.5 s and an accumulated dosage of about 54.1 electrons / Å2, which was fractionated into 22 frames (113 ms exposure each).

Structure determination and refinement

Beam-induced movement in each movie was corrected using MotionCor2[64] with dose-weighting applied, and the pixels were binned 2X to 0.833 Å size to generate a set of dose-weighted and another of dose-unweighted micrographs. Gctf[65] was used to estimate CTF (contrast transfer function) parameters on dose-unweighted micrographs. For both Cy5 and AP data (Extended Data Fig. 2 and 3), particles were picked automatically using Gautomatch (www2.mrc-lmb.cam.ac.uk/research/locally-developed-software/zhang-software/) as well as the build-in algorithms in RELION (ver. 3.1)[66] and extracted from dose-weighted micrographs, separately. To speed up the data processing, all particles were downscaled to a pixel size of 1.666 Å and a box size of 192×192. 2D and 3D classification were carried out in RELION (ver. 3.1)[66].

The C7CD non-duplicated particles were selected, combined and auto-refined. To improve the conformational purity of Core7 and XPC-DNA portions, subtraction and focused 3D classification without alignment were applied. Selected subsets of particles were re-extracted with a pixel size of 0.833 Å and a box size of 394×394, and subjected to CTF refinement, Bayesian polishing and auto-refinement in RELION (ver. 3.1). The final reconstruction of XPC_DNA reaches 4.2 Å resolution, and Core7 3.9 Å (Extended Data Fig. 2). The composite map of C7CD was generated by aligning and merging the two focused maps in UCSF Chimera[67]. For AP_C7CD, two different conformations were identified are refined to lower resolutions than Cy5_C7CD (Extended Data Fig. 3).

The C7CAD non-duplicated particles were selected, combined and subjected to auto-refinement. Volume subtraction was made on the XPC-DNA region, and the subtracted particles were subjected to focused 3D classification to segregate C7AD from C7CAD. To improve the density around the lesion-mimic Cy5 region in C7CAD, the XPC-DNA region was masked and treated with another round of focused 3D classification. Volume subtraction and 3D classification were also performed on the XPC C-term helices and their interacting Core7 region. The good classes belonging to C7AD, C7CAD and the XPC C-term helices were selected, and the corresponding raw particles were separately subjected to CTF refinement, Bayesian polishing and auto-refinement in RELION (ver. 3.1). Finally, the reconstruction of the Cy5-DNA C7CAD complexes reached 4.4 Å, C7AD 3.3 Å, and

the XPC C-term helical region 3.4 Å (Extended Data Fig. 2). For AP_C7CAD, similar data processing strategies were applied. C7CAD and the C7AD were also segregated and separately refined (Extended Data Fig. 3). DeepEMhancer was used to reduce noise in final maps[68].

Coordinates of TFIIH Core7 subunits (PDB: 6NMI and 6RO4) were used as initial models to build the C7CD and C7CAD/C7AD, respectively. Coordinates of XPA (PDB: 6RO4), CETN2 (PDB: 2GGM) and yeast XPC were initially docked into cryoEM maps. The human XPC structure predicted by AlphaFold (Q01831, residues 167–940)[27,69] was docked to replace yeast XPC. After initial fitting into the map using Chimera[67], all parts were manually adjusted in Coot[70]. A standard B-form DNA was fitted into the Cy5 cryoEM maps of C7CD and C7CAD initially, and the DNA surrounding the Cy5 was rebuilt manually in Coot. The AP-DNA in C7CD and C7CAD were built based on the Cy5-DNA. The final models were refined in PHENIX Real-Space refinement[71], and validated using MolProbity[72] (Extended Data Table 1). The “Gold Standard” of 0.143 Fourier Shell Correlation (FSC) criterion was applied to determine the resolution of each structure[73], and ResMap was used for the local resolution estimation[74] (Extended Data Fig. 2 and 3). Structural figures were made using Pymol version 2 (Pymol.org/2/).

EMSA Assay

Bot56 DNA was 5’-labeled using γ^{32}P-ATP and annealed with complementary Top strand (Top56_iCy5 or Top56_2AP). The lesion-containing DNA duplexes (5 to 10 nM) were incubated with equal molar of XPC, 5 to 20 nM XPA, Core7, and additional protein mixes in 20 μl reaction buffer (25 mM Tris-HCl (pH 8.0), 100 mM KCl, 8% Glycerol, 0.007% Triton X-100, 5 mM DTT, 7 mM $MgCl_2$) at 30°C for 30 min, and electrophoresed on a 4% native gel in 1X TBE buffer at 4°C. Gels were visualized using a Typhoon FLA 9500 Phosphor Imager. K_d of trimeric XPC and lesion DNA was estimated to be < 5 nM because 10 nM of each resulted in nearly full shift of labeled lesion DNA.

Dual Incision assay

The dual incision reaction was performed in 20 μl reaction buffer (25mM Tris-HCl pH 8.0, 50 mM KOAc, 5mM DTT, 0.1mg/ml BSA, 10% glycerol) with 0.025 μM substrate, 0.05 μM XPC, 0.1 μM Core7, 0.15 μM XPA, 0.1 μM RPA, 0.05 μM XPF, 0.025 μM XPG. Cy5-DNA and internally ^{32}P-labeled AP-DNA were generated by ligation of 2–3 oligos (Supplementary Table 1) and purified using denaturing PAGE. Dual incisions were monitored by fluorescence (Cy5) and radiation (AP). After pre-incubating NER proteins and DNA substrate at 37°C for 5 min, 3 mM ATP and 7.5 mM $MgCl_2$ were added to start the dual incision assay. The reactions at 37°C were stopped after 5, 10, 20, 30, or 60 min by adding proteinase K to 0.25 mg/ml and incubation at 56°C for 15min. To denature the DNA oligos, the reaction mixture was heated at 95°C for 5 min in TBE-Urea sample buffer (Bio-Rad) before loading onto 15% polyacrylamide TBE-Urea gel. The gel and dual incision products were scanned and analyzed by Typhoon FLA 9500 Phosphor Imager.

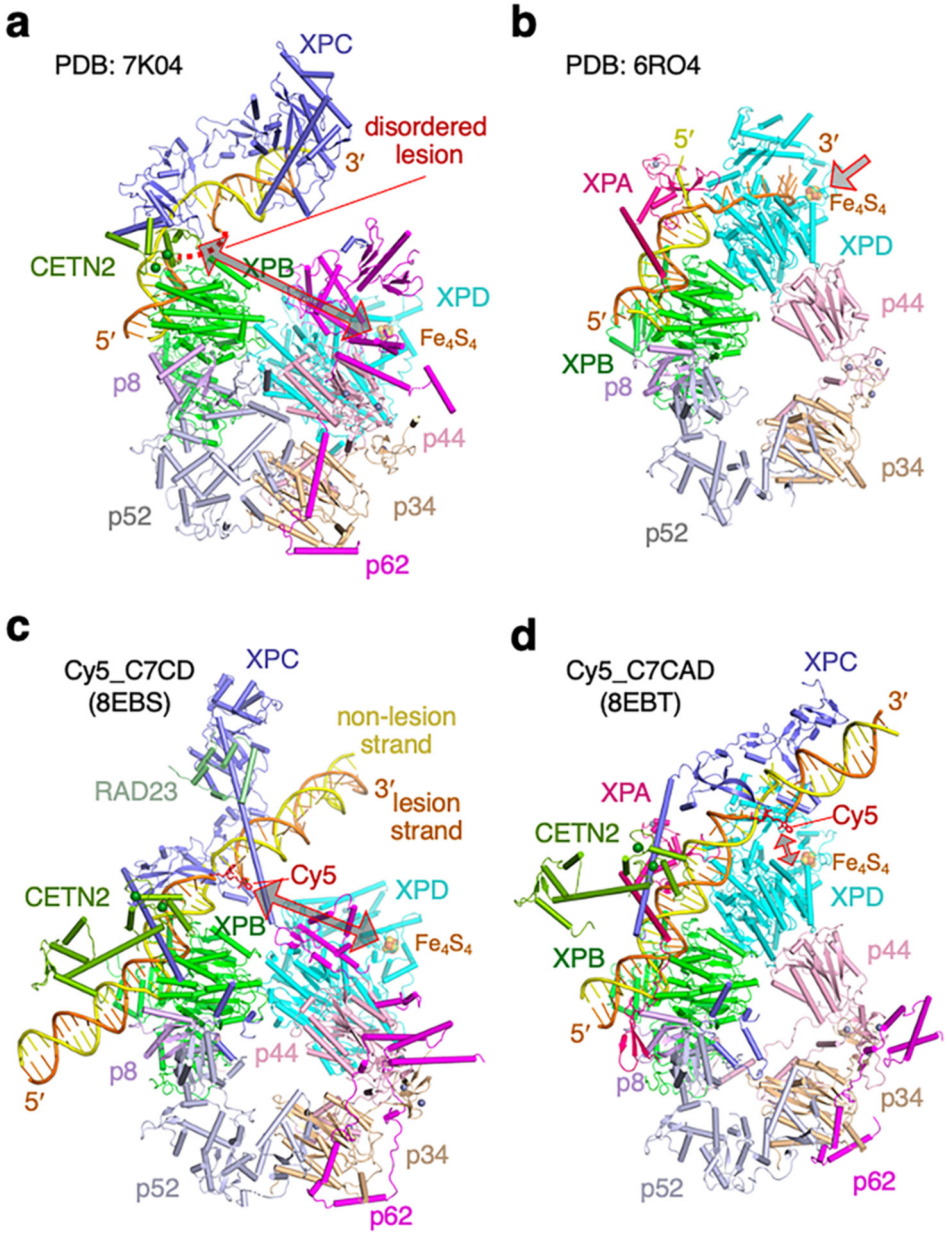
Extended Data

Extended Data Table 1.

CryoEM data collection and processing, and structural model refinement

	Cy5 complexes			AP complexes			
Data collection							
Pixel Size (Å)	1.245			1.245			
Defocus range(μm)	−2 to −3			−2 to −3			
Voltage (kV)	300			300			
Total electron dose (e^{-}/ Å^2)	54.12			54.12			
Extracted particles	1,645,921			1,145,832			
Type of complex	C7CD	C7CAD	C7AD	C7CD-conf1	C7CD-conf2	C7CAD	C7AD
PDB / EMDB accession code, focused map EMDB accession code	8EBS / 27996 EMD-29764	8EBT / 27997 EMD-28591	8EBU / 27998 EMD-29763	8EBV / 27999	8EBW / 28000	8EBX / 28001	8EBY / 28002.
Particles in final reconstruction / particles in composite map	188,623 / 89,824	127,485 / 45,955	169,304 / 177,287	65,301	66,841	103,713	137,438
Symmetry imposed	C1	C1	C1	C1	C1	C1	C1
Map resolution (Å, FSC=0.143)	3.8	3.9	3.3	7.1	5.6	3.6	3.6
Map sharpening B factor (Å^2)	−50	−50	−50	−50	−50	−50	−50
Model refinement and validation							
CC_mask	0.79	0.75	0.84	0.69	0.72	0.78	0.8
CC_volume	0.77	0.74	0.82	0.66	0.71	0.77	0.78
CC_peaks	0.7	0.64	0.73	0.57	0.61	0.68	0.67
Model resolution (Å, FSC=0.5)	4.0	4.0	3.6	8.2	7.0	3.7	3.7
Average B-factor (Å^2)							
Protein	133.2	105.4	155.9	524.2	421.3	145.6	135.1
Nucleotide	288.7	243.1	162.5	658.4	593.4	260.9	159.4
Ligand	162.8	298.6	285.2	539.5	451.0	349.3	246.5
RMSD Bond length (Å) / angle (°)	0.003 / 0.48	0.002 / 0.44	0.002 / 0.43	0.002 / 0.41	0.002 / 0.43	0.002 / 0.47	0.002 / 0.46
Ramachandran							
Most favored (%)	97.3	96.3	96.3	98.14	97.78	95.82	95.29
Allowed (%)	2.7	3.7	3.7	1.86	2.22	4.12	4.67
Disallowed (%)	0	0	0	0	0	0.06	0.04
Clash score	9.4	7.7	7.5	8.7	11.4	7.8	8.5

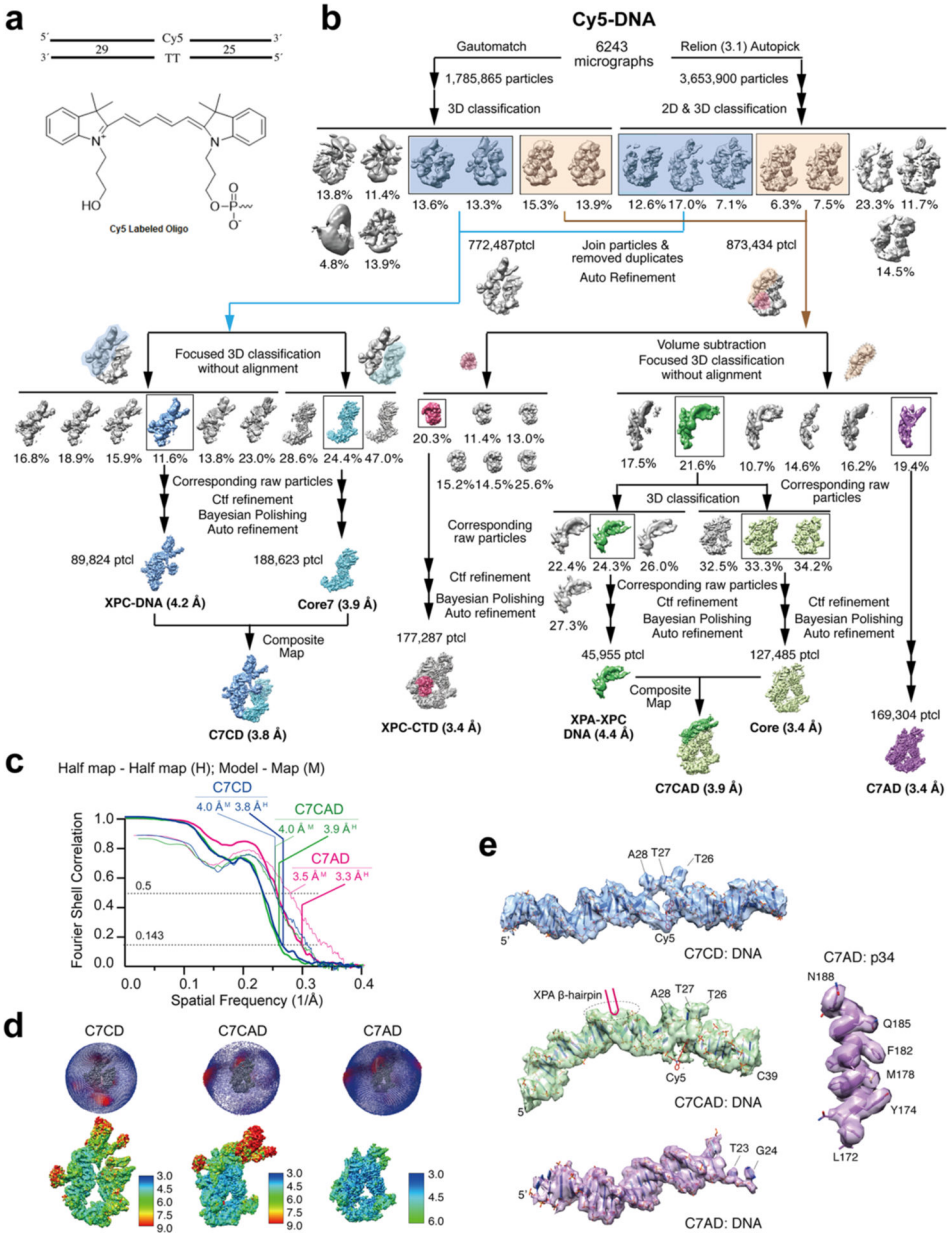
	Cy5 complexes			AP complexes			
Molprobity score	2.1	1.9	1.9	1.7	1.9	2.1	2.1
Rotamer outlier (%)	3.6	2.2	2.2	2.3	2.2	3.0	2.5
C-beta outliers (%)	0	0	0	0	0	0	0



Extended Data Fig. 1.

Structures of NER complexes. (**a**) Yeast Rad4 (XPC) complexed with Core7 and damaged DNA (orange and yellow) (PDB: 7K04 at 9.25 Å). (**b**) Human XPA and Core7 are complexed with undamaged but branched DNA (PDB: 6RO4). These structures are

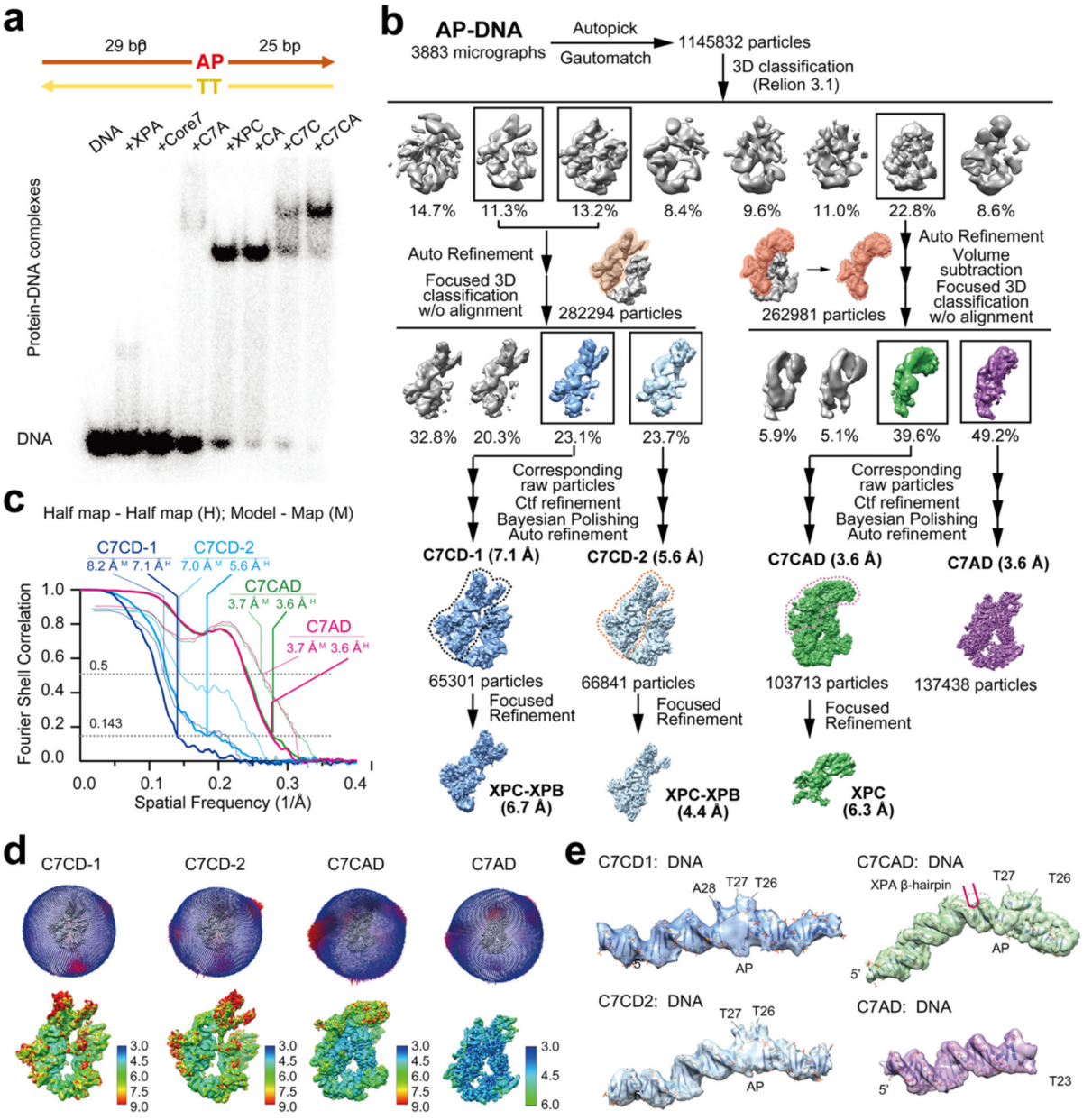
superimposed at XPB. The DNA damage site is far away and upstream of the lesion sensor Fe4S4 (marked by the grey arrowheads). (**c**) For comparison, the structures reported here, human XPC and Core7 complex with Cy5-DNA (C7CD), is shown after superposition with 7K04. (**d**) In human XPC, XPA and Core7 complexed with Cy5-DNA (C7CAD), the DNA lesion (Cy5) is downstream of the XPD motor (5¢ to 3¢) and close to the lesion sensor Fe4S4 of XPD. when XPD translocates along the lesion strand (orange), Cy5 would be "seen" and stall the XPD motor



Extended Data Fig. 2.

Structure determination of three Cy5 structures. (**a**) Diagram of Cy5-DNA and Cy5. (**b**) The workflow of cryoEM data processing and model generation. (**c**) FSC analysis of the quality

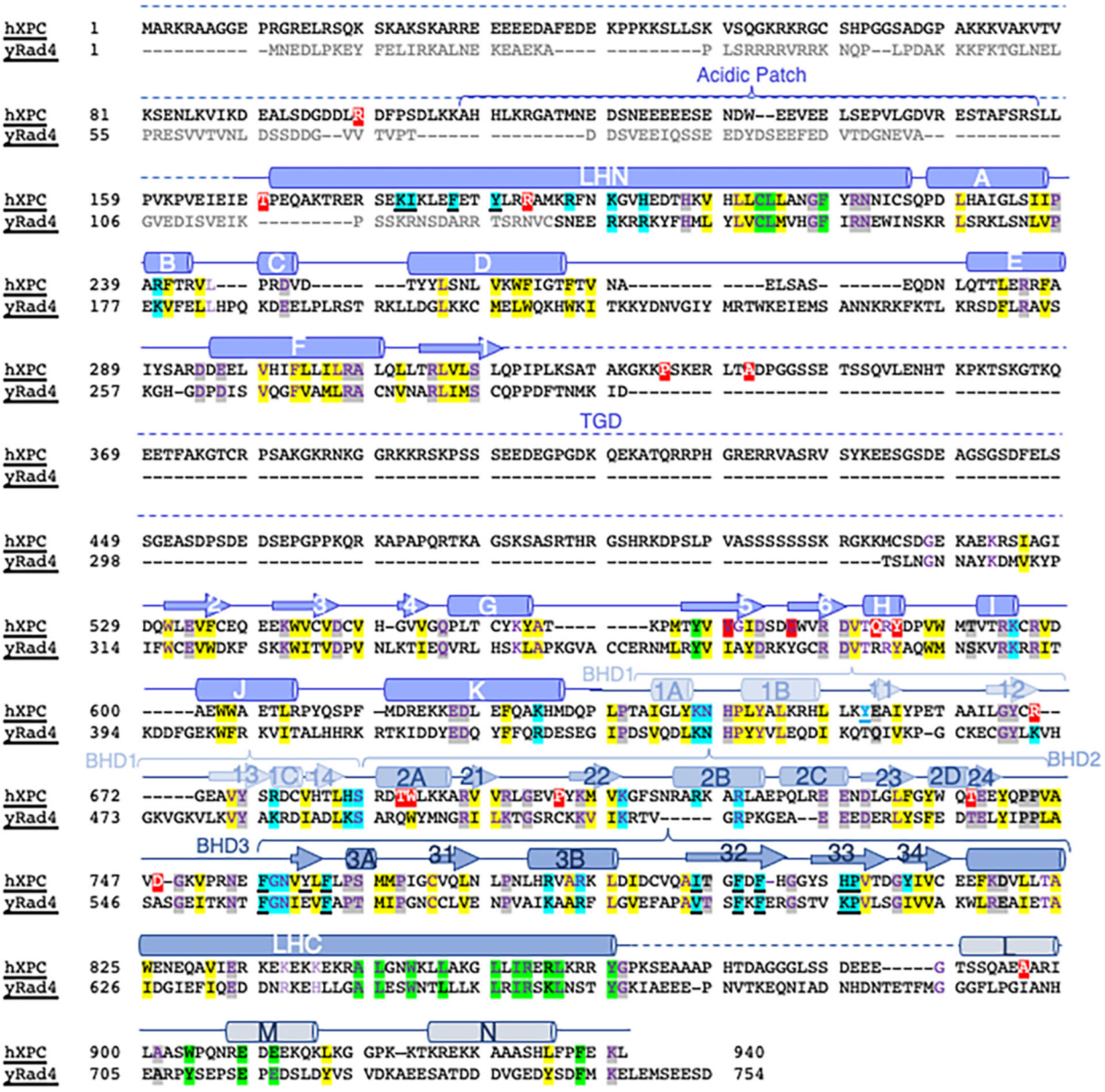
and map resolution and model fit of each complex structure. (**d**) For each complex, angular distributions of particles used for the final threedimensional reconstruction, and a surface presentation of its map colored according to the local resolution estimated by ResMap with the scale bar on the side, are shown. (**e**) Representative regions of the three cryoEM maps are superimposed with the finalstructural models



Extended Data Fig. 3. Structure determination of four AP structures.

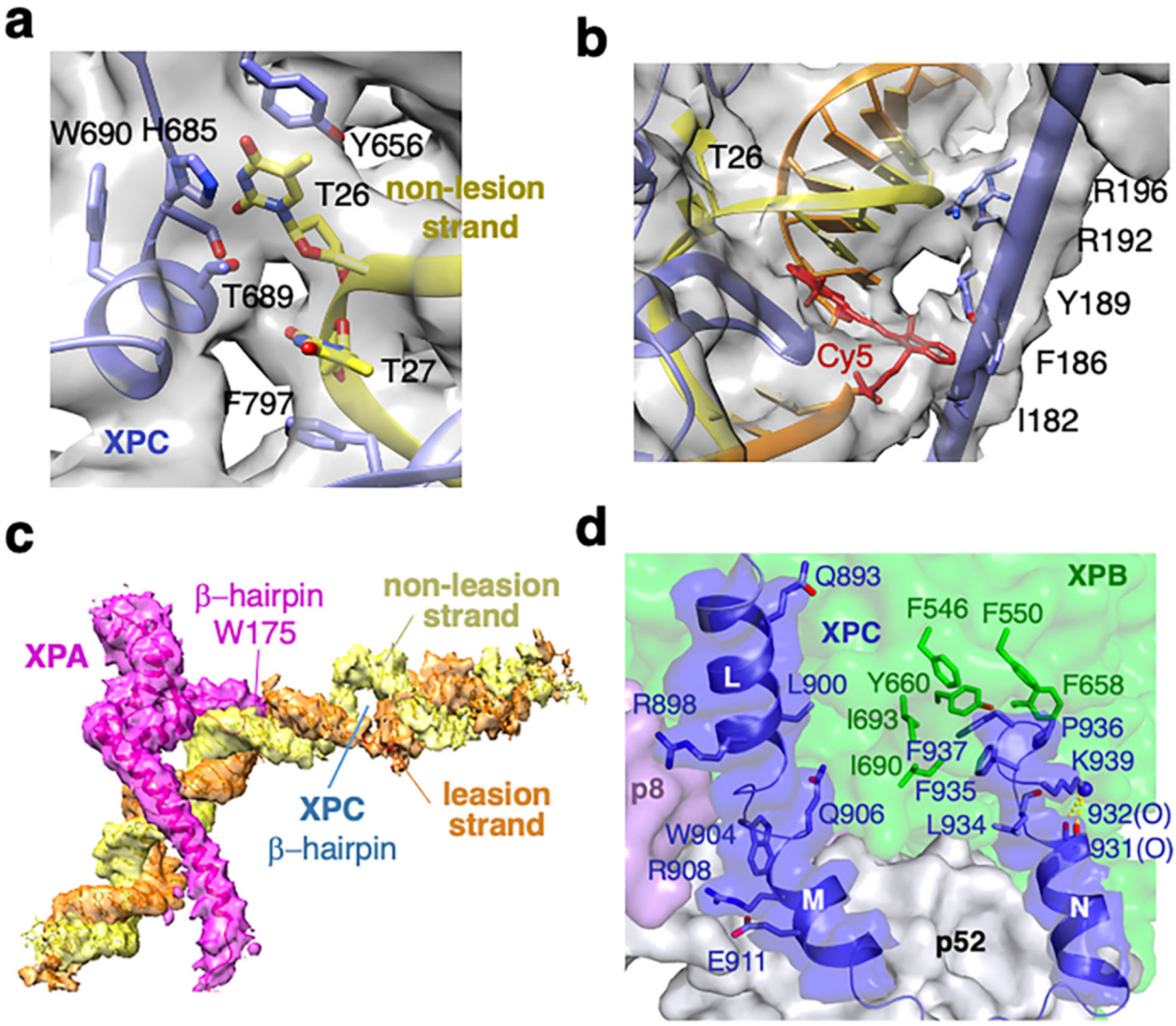
(**a**) Diagram of the AP-DNA, and EMSA results of 5 nM 32P-labeled AP-DNA binding by 5 nM each of XPA, Core7, Core7 and XPA (C7A), XPC, XPC and XPA (CA), Core7 and XPC (C7C) and Core7 with XPC and XPA (C7CA). The EMSA results were replicated at least six times. (**b**) The workflow of cryoEM data processing and model generation. (**c**)

FSC analysis of the quality and map resolution and model fit of each complex structure. (**d**) For each complex, angular distributions of particles used for the final three-dimensional reconstruction, and a surface presentation of its map, colored according to the local resolution estimated by ResMap with the scale bar on the side, are shown. (**e**) Representative regions of the three cryoEM maps (DNA) are superimposed with the final structural models. For gel source data of 3a, see Supplementary Figures 3.



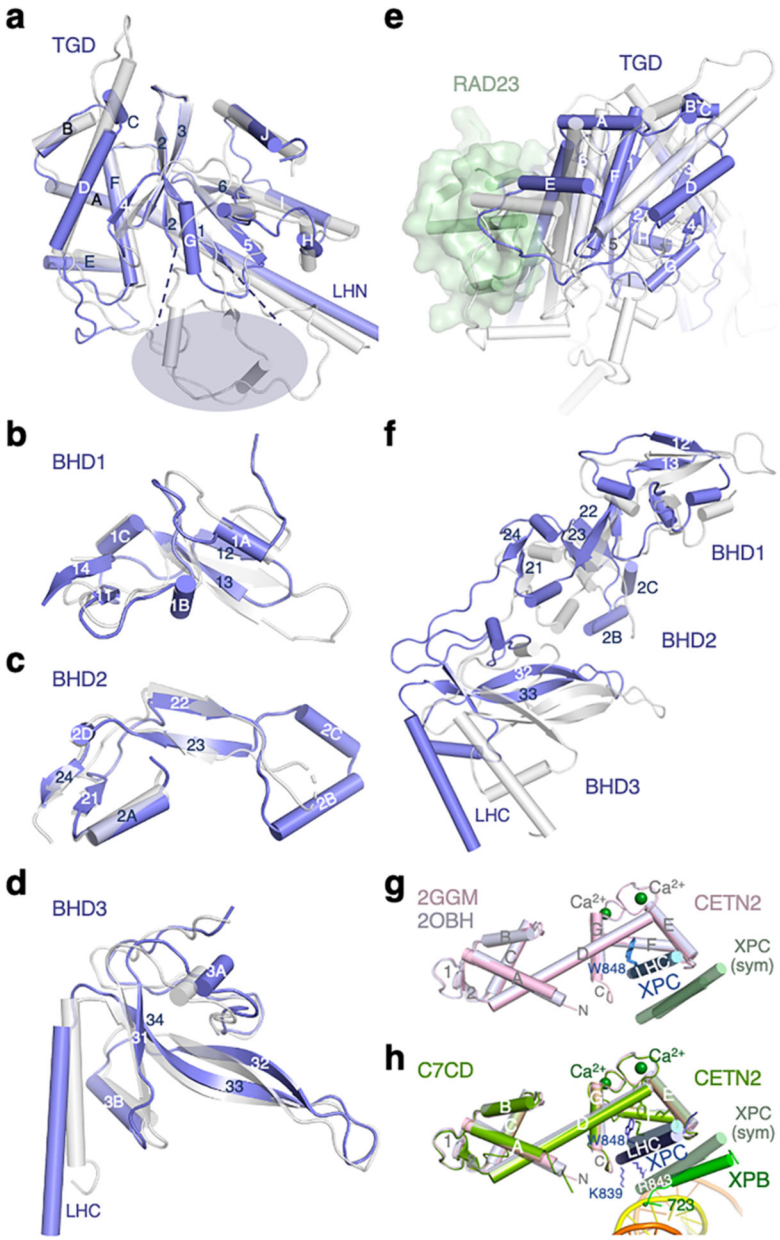
Extended Data Fig. 4. Structure-based sequence alignment of human XPC and yeast Rad4.
Conserved residues are highlighted in yellow (hydrophobic core), grey (structural stability), green (subunit interface), cyan (DNA binding, and underscore indicating base interactions), and red (disease mutation). Protein secondary structures are indicated by box (for helix) and arrow (strand). They are labeled alphabetically for helices and numerically for strands. In

BHD domains 1–3, secondary structures are preceded by domain name “1”, “2” and “3”. Disordered regions are indicated by dashed lines



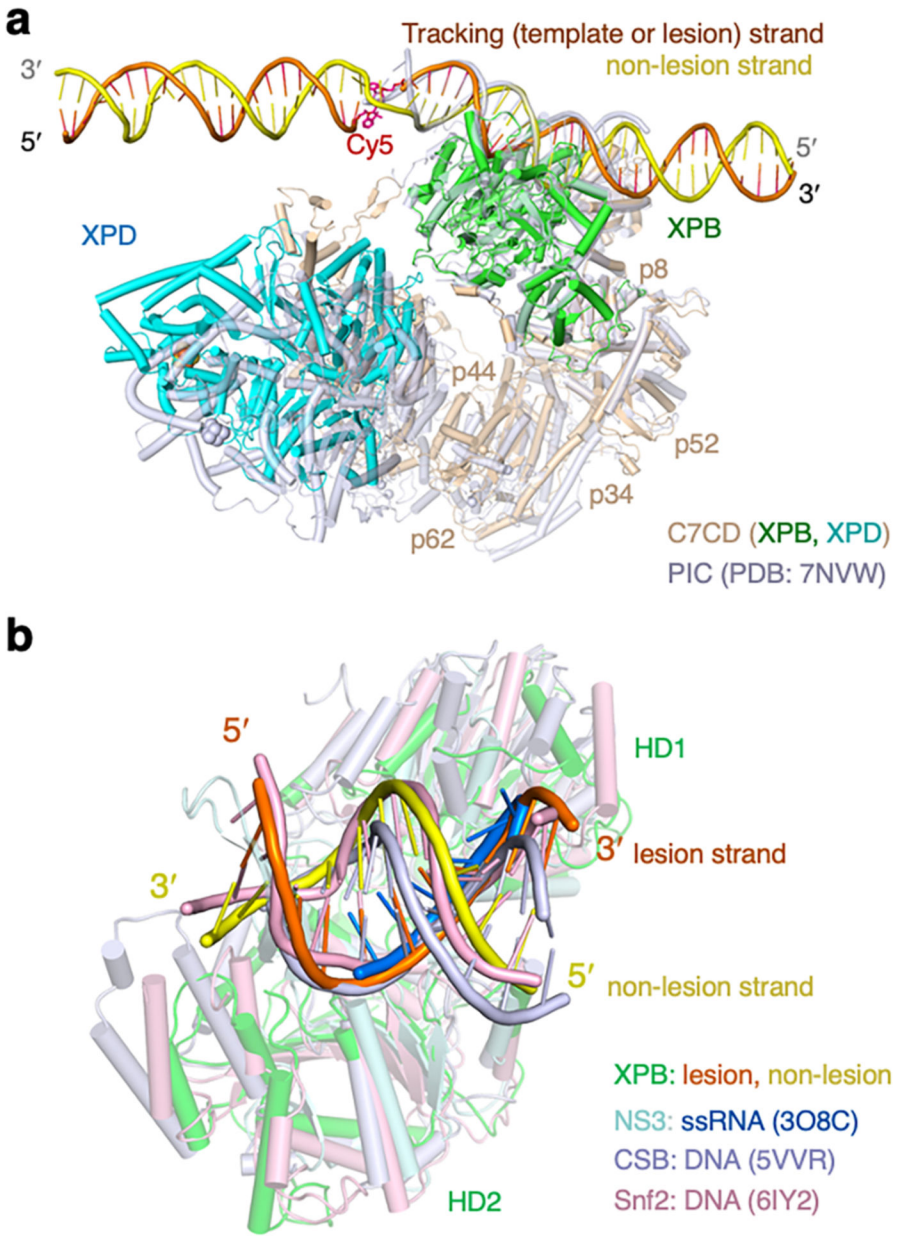
Extended Data Fig. 5.

cryoEM maps of DNA bound by XPC and XPA. (**a**) The flipped out T26 in Cy5_C7CD. (**b**) The LHN has close contacts with Cy5 and the non-lesion strand across the minor groove. The cryoEM map in the above two panels are shown as semitransparency? grey surface. (**c**) cryoEM map corresponding to XPA and DNA in Cy5_C7CAD. Map volume is color coded and labeled. (**d**) A close-up of the C-terminal 52 residues of XPC (aa 889–940). XPB, p52, and p8 of Core7 and XPC are represented by the cryoEM map of C7CAD and C7AD of Cy5-DNA and color coded. Helices L, M and N of XPC are show as ribbon cartoons and labeled. The penultimate K939 of XPC, which is shown in a stick model, caps the carboxyl end of helix N. Potential interactions between the sidechain amine of K939 (shown as a sphere) and carbonyl oxygens are indicated by dashed yellow lines. Residues F935, P936 and F937 of XPC are anchored in a hydrophobic pocket in XPB (green).



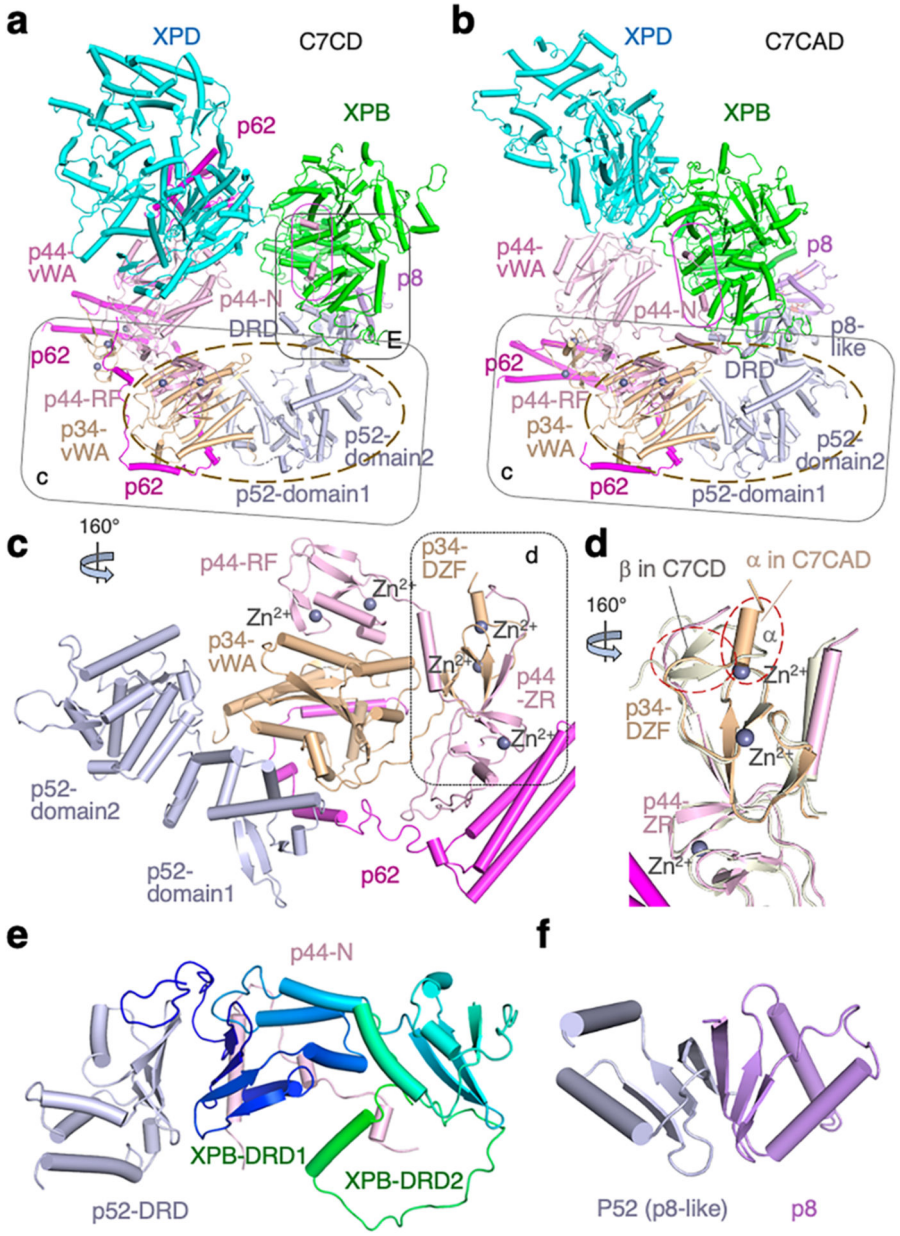
Extended Data Fig. 6. Domain comparison of XPC, Rad4, RAD23 and CETN2.

(**a**) Superposition of TGD of XPC (slate blue) and Rad4 (semi-transparent grey). (**b**) Superposition of BHD1 of XPC and Rad4. (**c**) Superposition of BHD2 of XPC and Rad4. (**d**) Superposition of BHD3 of XPC and Rad4. (**e**) Superposition of Rad23 and RAD23 (pale green cartoon with molecular surface) reveals that TGD domains of XPC (blue) and Rad4 (grey) differ by a 16° rotation. (**f**) Superposition of TGD domains of XPC and Rad4 reveals that BHD1, BHD2 and BHD3 diverge increasingly. (**g**) Crystal structures of CETN2 (2GGM in pink and 2OBH in light blue) complexed with XPC peptide (LHC, blue) are superimposed. Symmetry mate of XPC is shown in pale green. (**h**) CETN2 (light green) and XPB (dark green) in C7CD are included in superposition. The LHC (XPC, dark blue) is shifted and interacts with the C-terminal helix of XPB when complexed with Core7.



Extended Data Fig. 7.

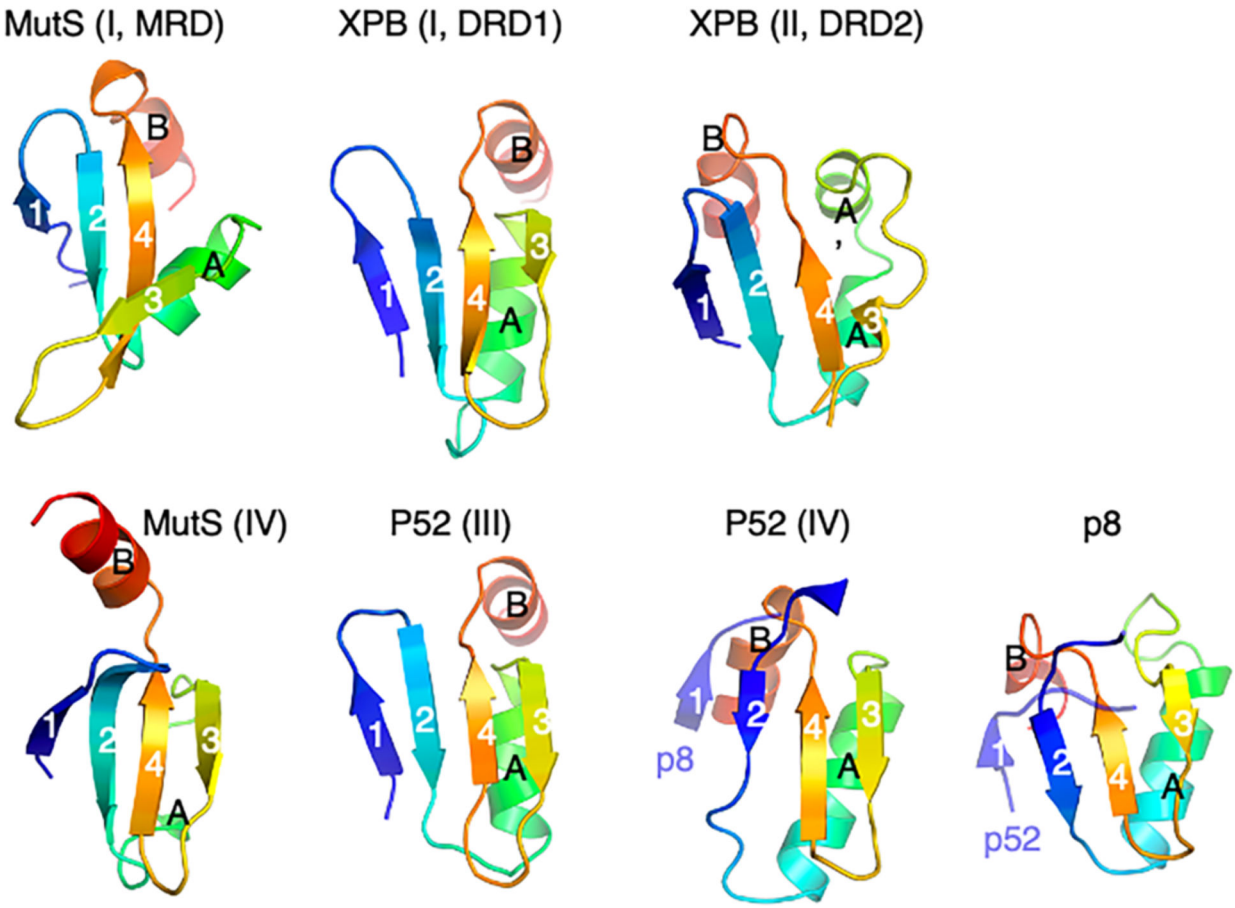
Structure comparison of C7CD with PIC and XPB with SF2 helicase. (**a**) Superposition of XPB (green) in C7CD and in human PIC (PDB: 7NVW, light grey) shows the bent DNA associated with XPB and different position of XPD (cyan in C7CD and light grey in PIC) in the U-shaped Core7. (**b**) Superposition of HD2 of four SF helicases, XPB, Rad26 (CSB homolog), Snf2 and NS3 reveals that the tracking strands superimpose well in all cases



Extended Data Fig. 8. Repetitive and flexible structure of TFIIH (Core7).

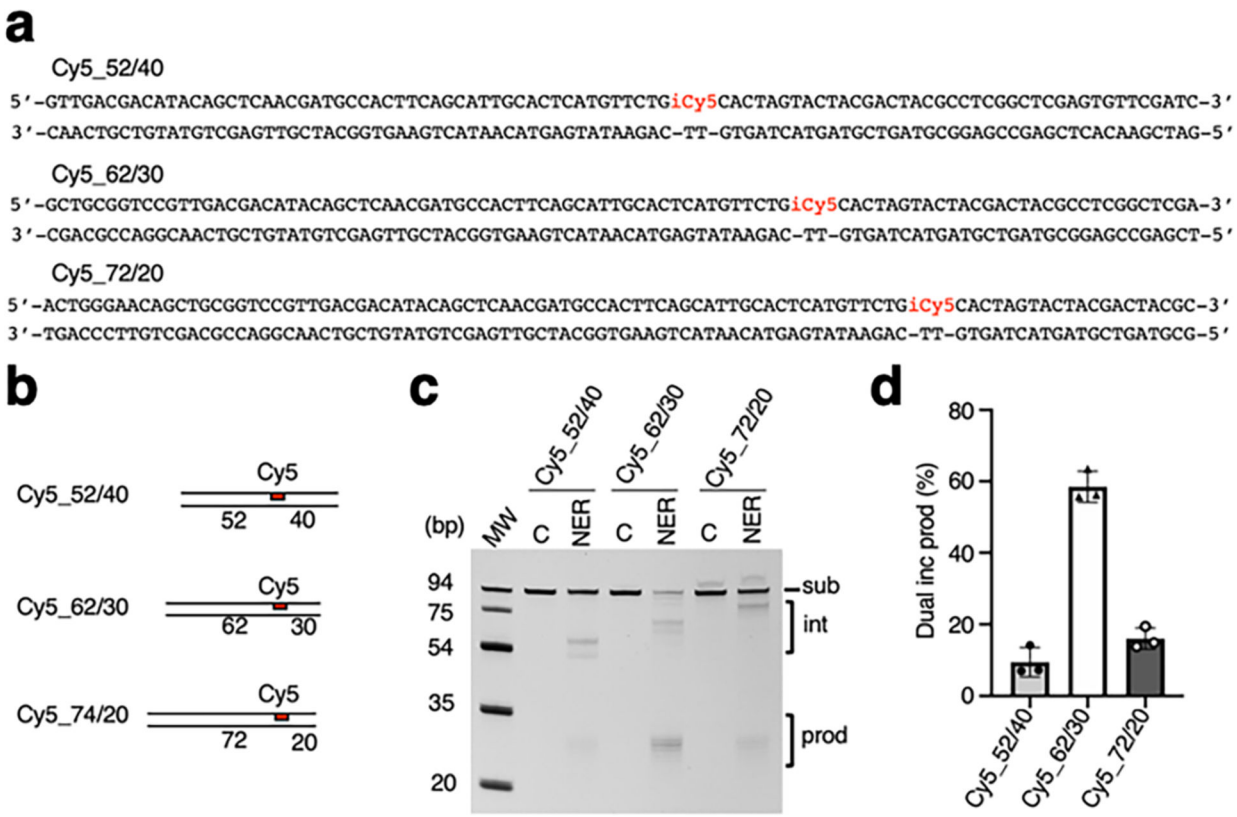
(**a**) The Ushaped Core7 in C7CD. The N-terminal helices of p44 that contact XPB are outlined in a rounded rectangle. The XPD (left) and XPB (right) arm are well separated. (**b**) The σ shaped Core7 in C7CAD with p34 superimposed to C7CD and viewed in the same orientation as in panel a. The interface at p34-p44 and p34-p52 (inside the dashed oval)? remain unchanged. (**c**) The stable interfaces of p34 with p44-RING finger (RF) and p52. The C-terminal p34-DZF (double Zinc finger) and p44-ZR (Zing Ribbon) domain are labeled. (**d**) A β hairpin of p34-DZF in C7CD is changed to a short α helix in C7CAD. A part of p62 becomes disordered in C7CAD. (**e**) The third domain of p52 (DRD fold) contacts the N-terminal DRD domain (blueish) of XPB, which is followed by the second DRD domain

(greenish) of XPB. The N-terminal helices of p44 (pink) contact the back side of XPB. (**f**) The fourth domain of p52 (grey) and p8 (light purple) form a heterodimer.



Extended Data Fig. 9. Comparison of DRD (Damage Recognition Domain) domains.

Two MutS DRDs (domains I and VI from 1EWQ) are shown on the left side for comparison. Five DRD domains in TFIIH are shown after superposition with MutS DRDs. Each DRD is colored in rainbow fashion from the blue N- to red C-terminus. Four β strands are labeled 1 to 4, and strands 2 and 4 are each followed by an α helix (A and B). In the P52-p8 heterodimer, the two subunits complement each other by supply the partner DRD with the first β strand (shown in semi-transparent blue and labeled 1').



Extended Data Fig. 10. Length of Cy5_DNA substrate required for efficient dual incision.

(**a**) Sequence of three Cy5 DNA substrates, each of which contains a total 94 bp but different upstream (left) and downstream length from Cy5 (right). (**b**) Diagrams of the three DNA substrates. (**c**) Dual incision results of each DNA substrate (sub) after incubation with Core7, XPC, XPA, RPA, XPF and XPG at 37°C for 60 min. DNA cleavage intermediate (int) and final product (prod) are marked. (**d**) Means and standard deviations (error bars) of triplicated dual incision reactions as well as individual data points are shown in the bar graph. For gel source data, see Supplementary Figures 5.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors thank Drs. R. Craigie, M. Gellert, H. Lans, D. Leahy and W. Vermeulen for critical reacting of the manuscript and J. Li for preparing carbon-filmed cryoEM grids. This work utilized the Cryo-Electron Microscopy Core facility, NIDDK, and the NIH Multi- Institute Cryo-EM Facility (MICEF). This research was supported by National Institute of Diabetes, Digestive and Kidney Disease (DK075037) to W.Y., and Grants-in-Aid (KAKENHI) (Grant Number JP16H06307 and JP21H03598) to K.S.

Data and Code Availability

The structures and cryoEM maps have been deposited with PDB and EMDB with accession codes of 8EBS, 8EBT and 8EBU, EMD-27996, 27997 and 27998 for C7CD, C7CAD and

C7AD of Cy5; 8EBV, 8EBW, 8EBX and 8EBY and EMD-27999, 28000,28001 and 28002 for C7CD1, C7CD2, C7CAD and C7AD of AP. The focused refinement maps of XPC-lesion DNA in Cy5_C7CD and the C-terminal domain of XPC in C7CAD and C7AD have been deposited with EMDB with accession codes EMD-29674 and 29673, respectively. These data will be released immediately upon publication. Other **r**esearch materials reported here are available upon request.

References

1. Hoeijmakers JH Nucleotide excision repair. II: From yeast to mammals. Trends in genetics : TIG 9, 211–217, doi:10.1016/0168-9525(93)90121-w (1993). [PubMed: 8337762]
2. Marteijn JA, Lans H, Vermeulen W & Hoeijmakers JH Understanding nucleotide excision repair and its roles in cancer and ageing. Nature reviews. Molecular cell biology 15, 465–481, doi:10.1038/nrm3822 (2014). [PubMed: 24954209]
3. Scharer OD Nucleotide excision repair in eukaryotes. Cold Spring Harb Perspect Biol 5, a012609, doi:10.1101/cshperspect.a012609 (2013).
4. Min JH & Pavletich NP Recognition of DNA damage by the Rad4 nucleotide excision repair protein. Nature 449, 570–575, doi:10.1038/nature06155 (2007). [PubMed: 17882165]
5. Kokic G et al. Structural basis of TFIIH activation for nucleotide excision repair. Nat Commun 10, 2885, doi:10.1038/s41467-019-10745-5 (2019). [PubMed: 31253769]
6. Aibara S, Schilbach S & Cramer P Structures of mammalian RNA polymerase II pre-initiation complexes. Nature 594, 124–128, doi:10.1038/s41586-021-03554-8 (2021). [PubMed: 33902107]
7. Schilbach S, Aibara S, Dienemann C, Grabbe F & Cramer P Structure of RNA polymerase II pre-initiation complex at 2.9 A defines initial DNA opening. Cell 184, 4064–4072 e4028, doi:10.1016/j.cell.2021.05.012 (2021). [PubMed: 34133942]
8. Yang C et al. Structural visualization of de novo transcription initiation by Saccharomyces cerevisiae RNA polymerase II. Molecular cell 82, 660–676 e669, doi:10.1016/j.molcel.2021.12.020 (2022). [PubMed: 35051353]
9. Schafer A et al. Functional and molecular genetic analyses of nine newly identified XPD-deficient patients reveal a novel mutation resulting in TTD as well as in XP/CS complex phenotypes. Experimental dermatology 22, 486–489, doi:10.1111/exd.12166 (2013). [PubMed: 23800062]
10. Liakos A, Lavigne MD & Fousteri M Nucleotide Excision Repair: From Neurodegeneration to Cancer. Adv Exp Med Biol 1007, 17–39, doi:10.1007/978-3-319-60733-7_2 (2017). [PubMed: 28840550]
11. Masutani C et al. Purification and cloning of a nucleotide excision repair complex involving the xeroderma pigmentosum group C protein and a human homologue of yeast RAD23. The EMBO journal 13, 1831–1843 (1994). [PubMed: 8168482]
12. Araki M et al. Centrosome protein centrin 2/caltractin 1 is part of the xeroderma pigmentosum group C complex that initiates global genome nucleotide excision repair. The Journal of biological chemistry 276, 18665–18672, doi:10.1074/jbc.M100855200 (2001). [PubMed: 11279143]
13. Sugasawa K et al. A multistep damage recognition mechanism for global genomic nucleotide excision repair. Genes & development 15, 507–521, doi:10.1101/gad.866301 (2001). [PubMed: 11238373]
14. Nishi R et al. Centrin 2 stimulates nucleotide excision repair by interacting with xeroderma pigmentosum group C protein. Mol Cell Biol 25, 5664–5674, doi:10.1128/MCB.25.13.5664-5674.2005 (2005). [PubMed: 15964821]
15. Paul D et al. Structure and mechanism of pyrimidine-pyrimidone (6–4) photoproduct recognition by the Rad4/XPC nucleotide excision repair complex. Nucleic acids research 47, 6015–6028, doi:10.1093/nar/gkz359 (2019). [PubMed: 31106376]
16. van Eeuwen T et al. Cryo-EM structure of TFIIH/Rad4-Rad23-Rad33 in damaged DNA opening in nucleotide excision repair. Nat Commun 12, 3338, doi:10.1038/s41467-021-23684-x (2021). [PubMed: 34099686]

17. Egly JM & Coin F A history of TFIIH: two decades of molecular biology on a pivotal transcription/repair factor. DNA repair 10, 714–721, doi:10.1016/j.dnarep.2011.04.021 (2011). [PubMed: 21592869]
18. Tirode F, Busso D, Coin F & Egly JM Reconstitution of the transcription factor TFIIH: assignment of functions for the three enzymatic subunits, XPB, XPD, and cdk7. Molecular cell 3, 87–95 (1999). [PubMed: 10024882]
19. Greber BJ, Toso DB, Fang J & Nogales E The complete structure of the human TFIIH core complex. Elife 8, doi:10.7554/eLife.44771 (2019).
20. Coin F et al. Nucleotide excision repair driven by the dissociation of CAK from TFIIH. Molecular cell 31, 9–20, doi:10.1016/j.molcel.2008.04.024 (2008). [PubMed: 18614043]
21. Naegeli H, Modrich P & Friedberg EC The DNA helicase activities of Rad3 protein of Saccharomyces cerevisiae and helicase II of Escherichia coli are differentially inhibited by covalent and noncovalent DNA modifications. The Journal of biological chemistry 268, 10386–10392 (1993). [PubMed: 8387518]
22. Sugasawa K, Akagi J, Nishi R, Iwai S & Hanaoka F Two-step recognition of DNA damage for mammalian nucleotide excision repair: Directional binding of the XPCcomplex and DNA strand scanning. Molecular cell 36, 642–653, doi:10.1016/j.molcel.2009.09.035 (2009). [PubMed: 19941824]
23. Li CL et al. Tripartite DNA Lesion Recognition and Verification by XPC, TFIIH, and XPA in Nucleotide Excision Repair. Molecular cell 59, 1025–1034, doi:10.1016/j.molcel.2015.08.012 (2015). [PubMed: 26384665]
24. Krasikova YS et al. Comparative analysis of interaction of human and yeast DNA damage recognition complexes with damaged DNA in nucleotide excision repair.The Journal of biological chemistry 288, 10936–10947, doi:10.1074/jbc.M112.444026 (2013). [PubMed: 23443653]
25. Kastner B et al. GraFix: sample preparation for single-particle electron cryomicroscopy. Nat Methods 5, 53–55, doi:10.1038/nmeth1139 (2008). [PubMed: 18157137]
26. Ribeiro-Silva C et al. Ubiquitin and TFIIH-stimulated DDB2 dissociation drives DNA damage handover in nucleotide excision repair. Nat Commun 11, 4868, doi:10.1038/s41467-020-18705-0 (2020). [PubMed: 32985517]
27. Jumper J et al. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589, doi:10.1038/s41586-021-03819-2 (2021). [PubMed: 34265844]
28. Kusakabe M et al. Histone deacetylation regulates nucleotide excision repair through an interaction with the XPC protein. iScience 25, 104040, doi:10.1016/j.isci.2022.104040 (2022). [PubMed: 35330687]
29. Sakai W et al. Functional impacts of the ubiquitin-proteasome system on DNA damage recognition in global genome nucleotide excision repair. Sci Rep 10, 19704, doi:10.1038/s41598-020-76898-2 (2020). [PubMed: 33184426]
30. Leongamornlert DA et al. Germline DNA Repair Gene Mutations in Young-onset Prostate Cancer Cases in the UK: Evidence for a More Extensive Genetic Panel. Eur Urol 76, 329–337, doi:10.1016/j.eururo.2019.01.050 (2019). [PubMed: 30777372]
31. Thompson JR, Ryan ZC, Salisbury JL & Kumar R The structure of the human centrin 2-xeroderma pigmentosum group C protein complex. The Journal of biological chemistry 281, 18746–18752, doi:10.1074/jbc.M513667200 (2006). [PubMed: 16627479]
32. Charbonnier JB et al. Structural, thermodynamic, and cellular characterization of human centrin 2 interaction with xeroderma pigmentosum group C protein. J Mol Biol 373, 1032–1046, doi:10.1016/j.jmb.2007.08.046 (2007). [PubMed: 17897675]
33. Lafrance-Vanasse J, Arseneault G, Cappadocia L, Legault P & Omichinski JG Structural and functional evidence that Rad4 competes with Rad2 for binding to the Tfb1 subunit of TFIIH in NER. Nucleic acids research 41, 2736–2745, doi:10.1093/nar/gks1321 (2013). [PubMed: 23295669]
34. Okuda M, Kinoshita M, Kakumu E, Sugasawa K & Nishimura Y Structural Insight into the Mechanism of TFIIH Recognition by the Acidic String of the Nucleotide Excision Repair Factor XPC. Structure 23, 1827–1837, doi:10.1016/j.str.2015.07.009 (2015). [PubMed: 26278177]

35. Matsuda T et al. DNA repair protein XPA binds replication protein A (RPA). The Journal of biological chemistry 270, 4152–4157, doi:10.1074/jbc.270.8.4152 (1995). [PubMed: 7876167]
36. Mer G et al. Structural basis for the recognition of DNA repair proteins UNG2, XPA, and RAD52 by replication factor RPA. Cell 103, 449–456 (2000). [PubMed: 11081631]
37. Ikegami T et al. Solution structure of the DNA- and RPA-binding domain of the human repair factor XPA. Nat Struct Biol 5, 701–706, doi:10.1038/1400 (1998). [PubMed: 9699634]
38. Koch SC et al. Structural insights into the recognition of cisplatin and AAF-dG lesion by Rad14 (XPA). Proceedings of the National Academy of Sciences of the United States of America 112, 8272–8277, doi:10.1073/pnas.1508509112 (2015). [PubMed: 26100901]
39. Simon N, Ebert C & Schneider S Structural Basis for Bulky-Adduct DNA-Lesion Recognition by the Nucleotide Excision Repair Protein Rad14. Chemistry 22, 10782–10785, doi:10.1002/chem.201602438 (2016). [PubMed: 27223336]
40. Satokata I, Tanaka K, Yuba S & Okada Y Identification of splicing mutations of the last nucleotides of exons, a nonsense mutation, and a missense mutation of the XPAC gene as causes of group A xeroderma pigmentosum. Mutation research 273, 203–212, doi:10.1016/0921-8777(92)90081-d (1992). [PubMed: 1372103]
41. Takahashi Y et al. XPA gene mutations resulting in subtle truncation of protein in xeroderma pigmentosum group A patients with mild skin symptoms. J Invest Dermatol 130, 2481–2488, doi:10.1038/jid.2010.137 (2010). [PubMed: 20574439]
42. Theil AF et al. Disruption of TTDA results in complete nucleotide excision repair deficiency and embryonic lethality. PLoS Genet 9, e1003431, doi:10.1371/journal.pgen.1003431 (2013).
43. Zadori D et al. Predominant neurological phenotype in a Hungarian family with two novel mutations in the XPA gene-case series. Neurol Sci 41, 125–129, doi:10.1007/s10072-019-04044-6 (2020). [PubMed: 31478152]
44. Nishi R, Sakai W, Tone D, Hanaoka F & Sugasawa K Structure-function analysis of the EF-hand protein centrin-2 for its intracellular localization and nucleotide excision repair. Nucleic acids research 41, 6917–6929, doi:10.1093/nar/gkt434 (2013). [PubMed: 23716636]
45. Huang KL et al. Pathogenic Germline Variants in 10,389 Adult Cancers. Cell 173, 355–370 e314, doi:10.1016/j.cell.2018.03.039 (2018). [PubMed: 29625052]
46. Sanyal S et al. Polymorphisms in DNA repair and metabolic genes in bladder cancer. Carcinogenesis 25, 729–734, doi:10.1093/carcin/bgh058 (2004). [PubMed: 14688016]
47. Fan L et al. Conserved XPB core structure and motifs for DNA unwinding: implications for pathway selection of transcription or excision repair. Molecular cell 22, 27–37, doi:10.1016/j.molcel.2006.02.017 (2006). [PubMed: 16600867]
48. Vitorino M et al. Solution structure and self-association properties of the p8 TFIIH subunit responsible for trichothiodystrophy. J Mol Biol 368, 473–480, doi:10.1016/j.jmb.2007.02.020 (2007). [PubMed: 17350038]
49. Aguilar-Fuentes J et al. p8/TTDA overexpression enhances UV-irradiation resistance and suppresses TFIIH mutations in a Drosophila trichothiodystrophy model. PLoS Genet 4, e1000253, doi:10.1371/journal.pgen.1000253 (2008).
50. Rudolf J, Makrantoni V, Ingledew WJ, Stark MJ & White MF The DNA repair helicases XPD and FancJ have essential iron-sulfur domains. Molecular cell 23, 801–808, doi:10.1016/j.molcel.2006.07.019 (2006). [PubMed: 16973432]
51. Fan L et al. XPD helicase structures and activities: insights into the cancer and aging phenotypes from XPD mutations. Cell 133, 789–800, doi:10.1016/j.cell.2008.04.030 (2008). [PubMed: 18510924]
52. Wolski SC et al. Crystal structure of the FeS cluster-containing nucleotide excision repair helicase XPD. PLoS biology 6, e149, doi:10.1371/journal.pbio.0060149 (2008). [PubMed: 18578568]
53. Xu J et al. Structural basis for the initiation of eukaryotic transcription-coupled DNA repair. Nature 551, 653–657, doi:10.1038/nature24658 (2017). [PubMed: 29168508]
54. Li M et al. Mechanism of DNA translocation underlying chromatin remodelling by Snf2. Nature 567, 409–413, doi:10.1038/s41586-019-1029-2 (2019). [PubMed: 30867599]
55. Appleby TC et al. Visualizing ATP-dependent RNA translocation by the NS3 helicase from HCV. J Mol Biol 405, 1139–1153, doi:10.1016/j.jmb.2010.11.034 (2011). [PubMed: 21145896]

56. Hu J, Adar S, Selby CP, Lieb JD & Sancar A Genome-wide analysis of human global and transcription-coupled excision repair of UV damage at single- nucleotide resolution. Genes & development 29, 948–960, doi:10.1101/gad.261271.115 (2015). [PubMed: 25934506]
57. Moggs JG, Yarema KJ, Essigmann JM & Wood RD Analysis of incision sites produced by human cell extracts and purified proteins during nucleotide excision repair of a 1,3-intrastrand d(GpTpG)-cisplatin adduct. The Journal of biological chemistry 271, 7177–7186, doi:10.1074/jbc.271.12.7177 (1996). [PubMed: 8636155]
58. Staresincic L et al. Coordination of dual incision and repair synthesis in human nucleotide excision repair. The EMBO journal 28, 1111–1120, doi:10.1038/emboj.2009.49 (2009). [PubMed: 19279666]
59. O'Donovan A, Davies AA, Moggs JG, West SC & Wood RD XPG endonuclease makes the 3' incision in human DNA nucleotide excision repair. Nature 371, 432–435, doi:10.1038/371432a0 (1994). [PubMed: 8090225]
60. Yasuda G et al. In vivo destabilization and functional defects of the xeroderma pigmentosum C protein caused by a pathogenic missense mutation. Mol Cell Biol 27, 6606–6614, doi:10.1128/MCB.02166-06 (2007). [PubMed: 17682058]
61. Henricksen LA, Umbricht CB & Wold MS Recombinant replication protein A: expression, complex formation, and functional characterization. The Journal of biological chemistry 269, 11121–11132 (1994). [PubMed: 8157639]
62. Kim M-S, Lapkouski M, Yang W & Gellert M Crystal structure of the V(D)J recombinase RAG1-RAG2. Nature 518, 507–511, doi:10.1038/nature14174 (2015). [PubMed: 25707801]
63. Schorb M, Haberbosch I, Hagen WJH, Schwab Y & Mastronarde DN Software tools for automated transmission electron microscopy. Nat Methods 16, 471–477, doi:10.1038/s41592-019-0396-9 (2019). [PubMed: 31086343]
64. Zheng SQ, Palovcak E, Armache JP, Verba KA, Cheng Y, and Agard DA MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nat Methods 14, 331–332 (2017). [PubMed: 28250466]
65. Zhang K Gctf: Real-time CTF determination and correction. J Struct Biol 193, 1–12, doi:10.1016/j.jsb.2015.11.003 (2016). [PubMed: 26592709]
66. Fernandez-Leiro R, and Scheres SHW A pipeline approach to single-particle processing in RELION. Acta Crystallogr D Struct Biol 73, 496–502 (2017). [PubMed: 28580911]
67. Pettersen EF et al. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem 25, 1605–1612, doi:10.1002/jcc.20084 (2004). [PubMed: 15264254]
68. Sanchez-Garcia R et al. DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. Commun Biol 4, 874, doi:10.1038/s42003-021-02399-1 (2021). [PubMed: 34267316]
69. Varadi M et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic acids research 50, D439–D444, doi:10.1093/nar/gkab1061 (2022). [PubMed: 34791371]
70. Emsley P, Lohkamp B, Scott WG, and Cowtan K Features and development of Coot. Acta Crystallogr D Biol Crystallogr 66, 486–501 (2010). [PubMed: 20383002]
71. Adams PD et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 66, 213–221, doi:S0907444909052925 [pii] 10.1107/S0907444909052925 (2010). [PubMed: 20124702]
72. Chen VB et al. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr 66, 12–21, doi:10.1107/S0907444909042073 (2010). [PubMed: 20057044]
73. Swint-Kruse L, and Brown CS Resmap: automated representation of macromolecular interfaces as two-dimensional networks. Bioinformatics 21, 3327–3328 (2005). [PubMed: 15914544]
74. Kucukelbir A, Sigworth FJ & Tagare HD Quantifying the local resolution of cryo-EM density maps. Nat Methods 11, 63–65, doi:10.1038/nmeth.2727 (2014). [PubMed: 24213166]

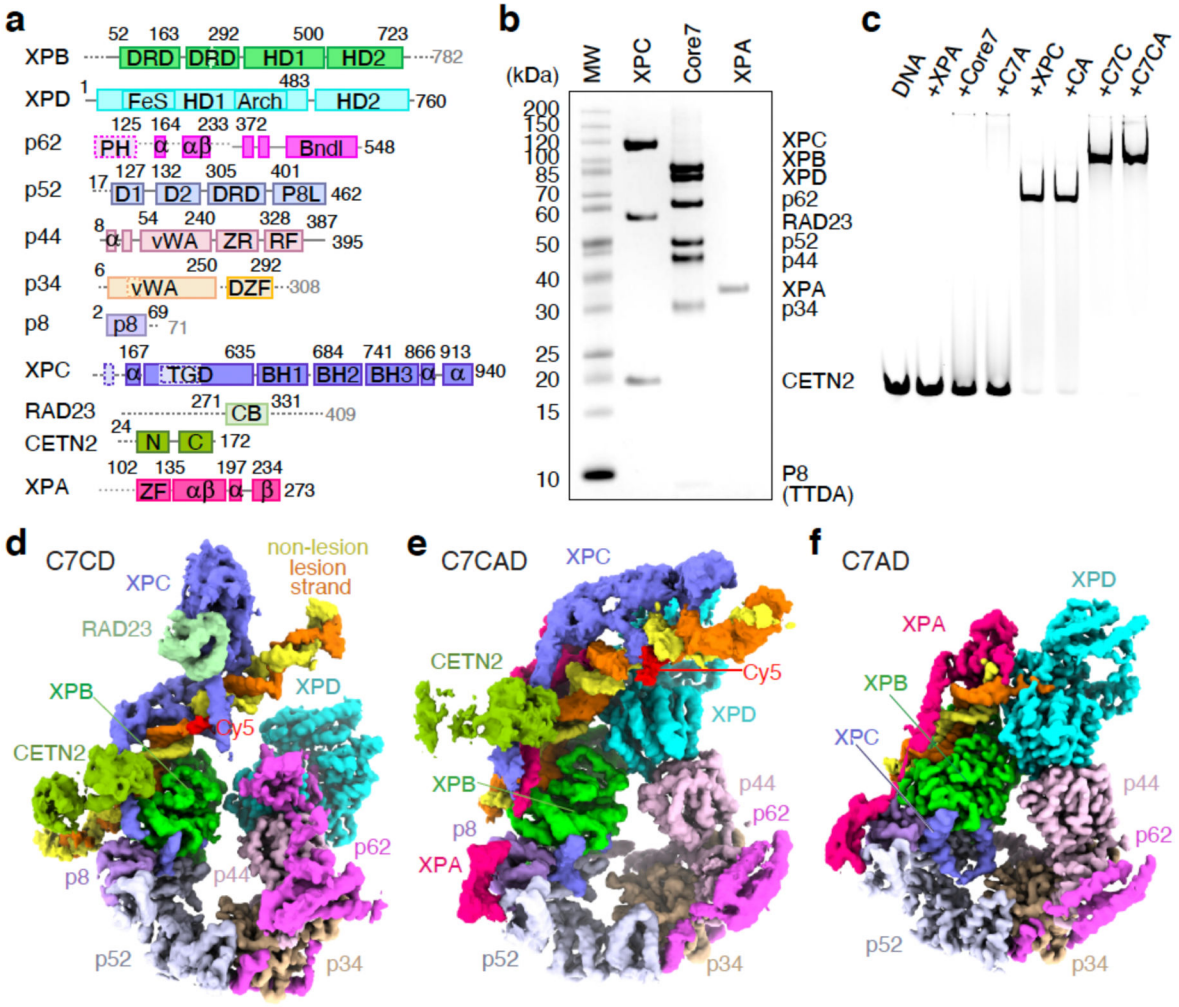


Fig. 1. Structures of lesion recognition and handoff.

(**a**) Primary structure of TFIIH (Core7), XPC and XPA. Structural domains are shown as rectangles and annotated. Each protein has a distinct color. Dotted lines and rectangles indicate disordered regions. (**b**) A protein SDS gel shows the purity of XPC, XPA and Core7. (**c**) EMSA results of 10 nM Cy5-DNA binding by 10 nM XPC, XPA, Core7, Core7 and XPA (C7A), XPC and XPA (CA), Core7 and XPC (C7C), and Core7 with XPC and XPA (C7CA). Data shown in 1b and 1c were reproduced at least six times. (**d-f**) cryoEM maps of C7CD (d), C7CAD (e) and C7AD (f) are color coded as in Fig. 1a. For gel source data of 1b and 1c, see Supplementary Figures 1 and 2.

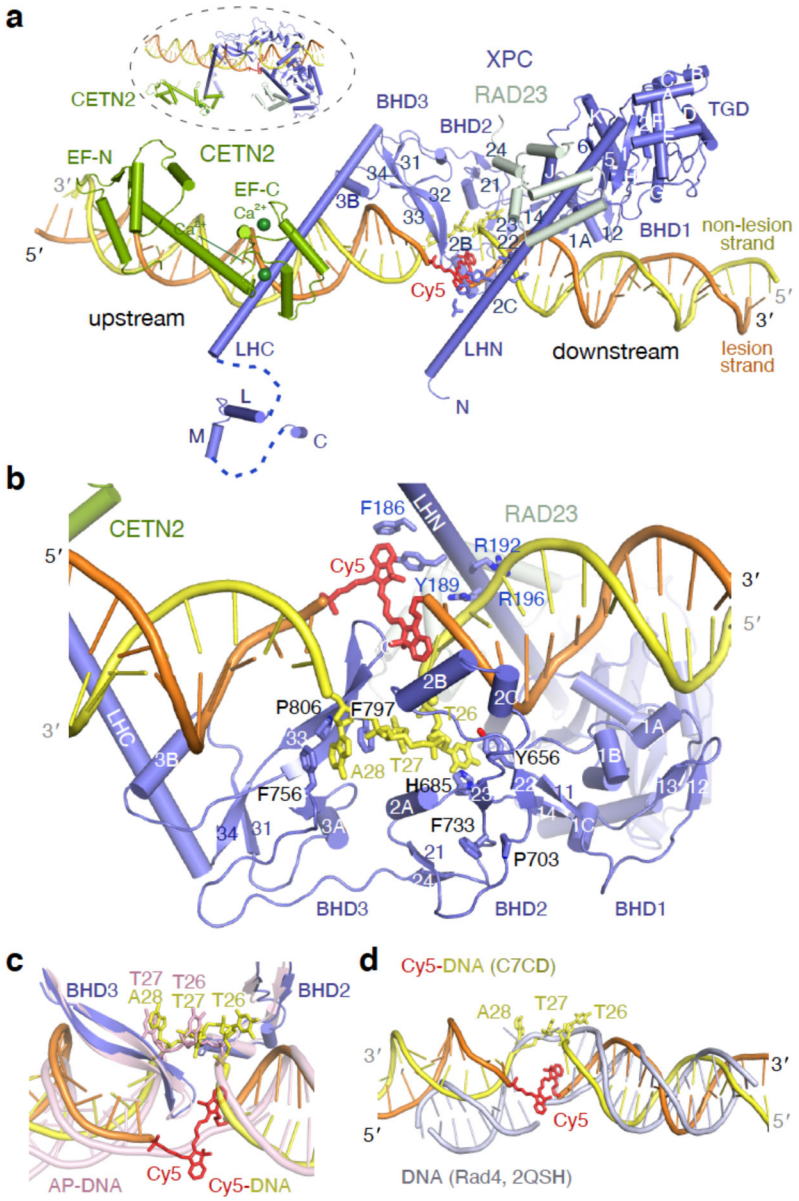


Fig. 2. Structure of human XPC-DNA complex.

(**a**) DNA duplex (yellow non-lesion and orange lesion strand) is bound by XPC (slate blue). RAD23 (pale green) and CETN2 (pea green) stabilize XPC. An orthogonal view is shown in the insert. (**b**) Lesion is recognized by BHD2 and BHD3. The flipped out Cy5 and opposite normal bases are stabilized by aromatic sidechains. (**c**) The flipped-out bases in Cy5 (multicolor) and AP-DNA (Conf2, pink) are shifted by 1 bp but overlap. (**d**) Overlay of the Cy5-DNA (yellow and orange) complexed XPC and DNA bound by Rad4 (PDB: 2QSH). DNA lesions complexed with Rad4 are disordered.

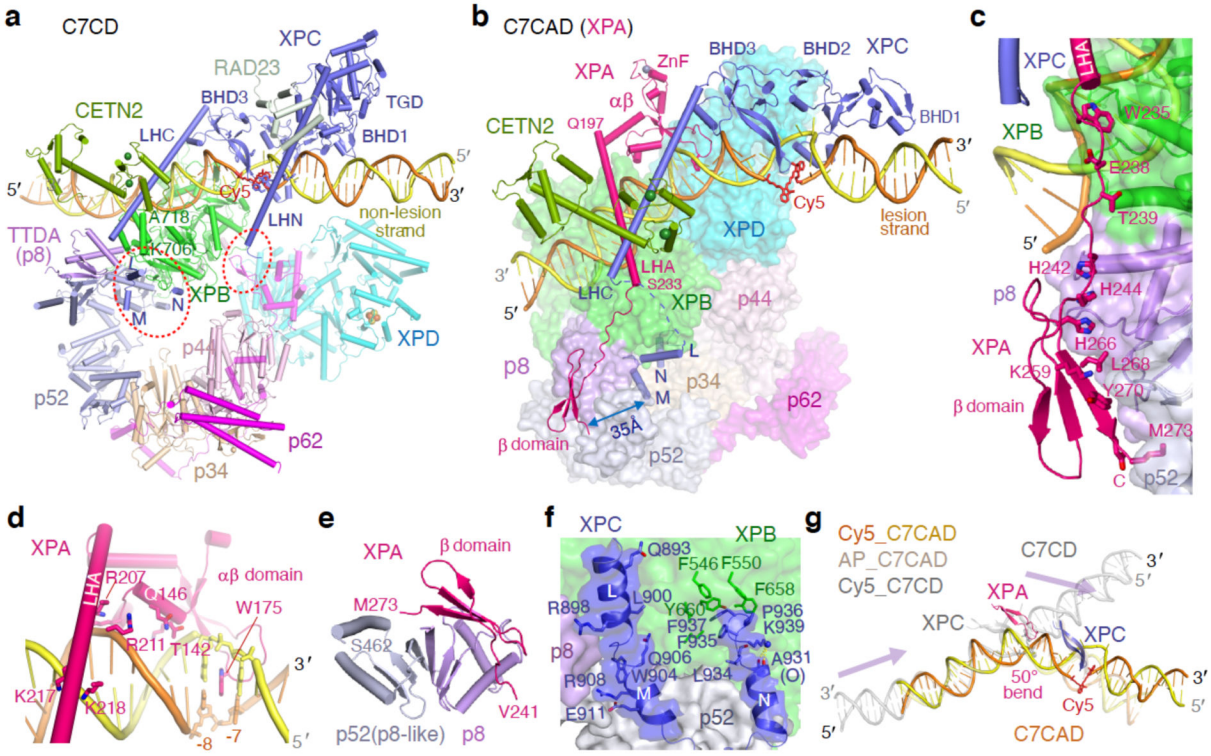


Fig. 3. Structures of C7CD and C7CAD.

(**a**) In C7CD, where XPC and Core7 interact are circled in red. (**b**) In C7CAD, Core7 is shown in molecular surface, and others in ribbon diagrams. (**c**) XPA contacts XPB, p8 and p52 sequentially. (**d**) The αβ domain of XPA binds and kinks the DNA duplex. (**e**) p52 and p8 form a heterodimer, and p8 binds XPA. (**f**) L, M and N helices of XPC are anchored in Core7. The cryoEM map (C7CAD) is shown as semi-transparent surface. (**g**) DNA in C7CD (light grey), Cy5_ (orange/ yellow) and AP_C7CAD (wheat) are shown after superposition of XPB. The β hairpins of XPA (hot pink) and XPC (blue) mark the distorted DNA sites. Light purple arrows indicate the shift of DNA and XPC from C7CD (light grey) to C7CAD with XPB superimposed.

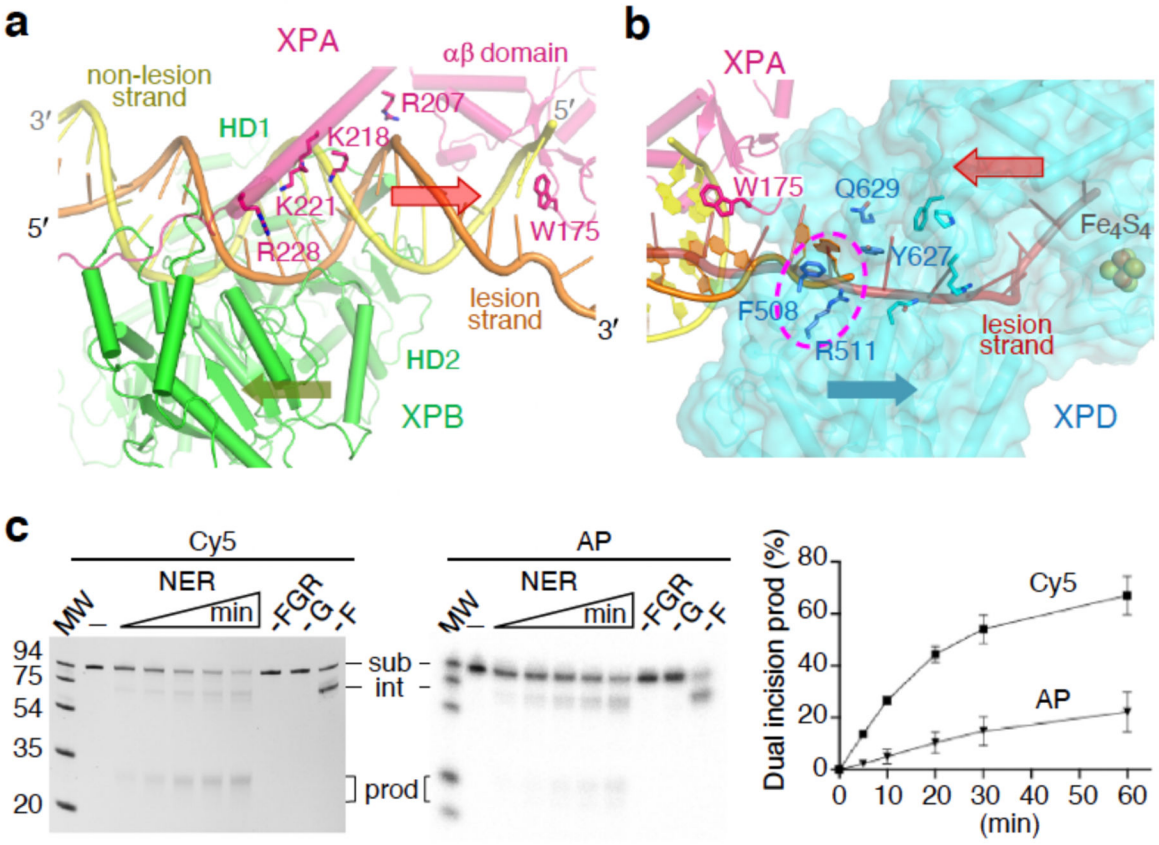


Fig. 4. DNA translocation and dual incision.

(**a**) XPB binds the upstream DNA opposite XPA and moves along the lesion strand (orange) 3′ to 5′ (indicated by the olive arrow). If the protein is stationary, DNA would move in the opposite direction (indicated by the red arrow). (**b**) Structure of C7AD with the XPD-bound lesion strand borrowed from 6RO4 (brown). The red arrow indicates the direction of ssDNA translocation by stationary XPD. (**c**) Dual incision of Cy5- and AP-DNA required Core7, XPC, XPA, XPF (F), XPG (G), RPA (R) and ATP. Bands of substrate (sub), intermediate (int) and product (prod) are indicated. Means and standard deviations (error bars) of triplicated dual incision reactions are shown in the plot. For gel source data of 4c, see Supplementary Figures 4.

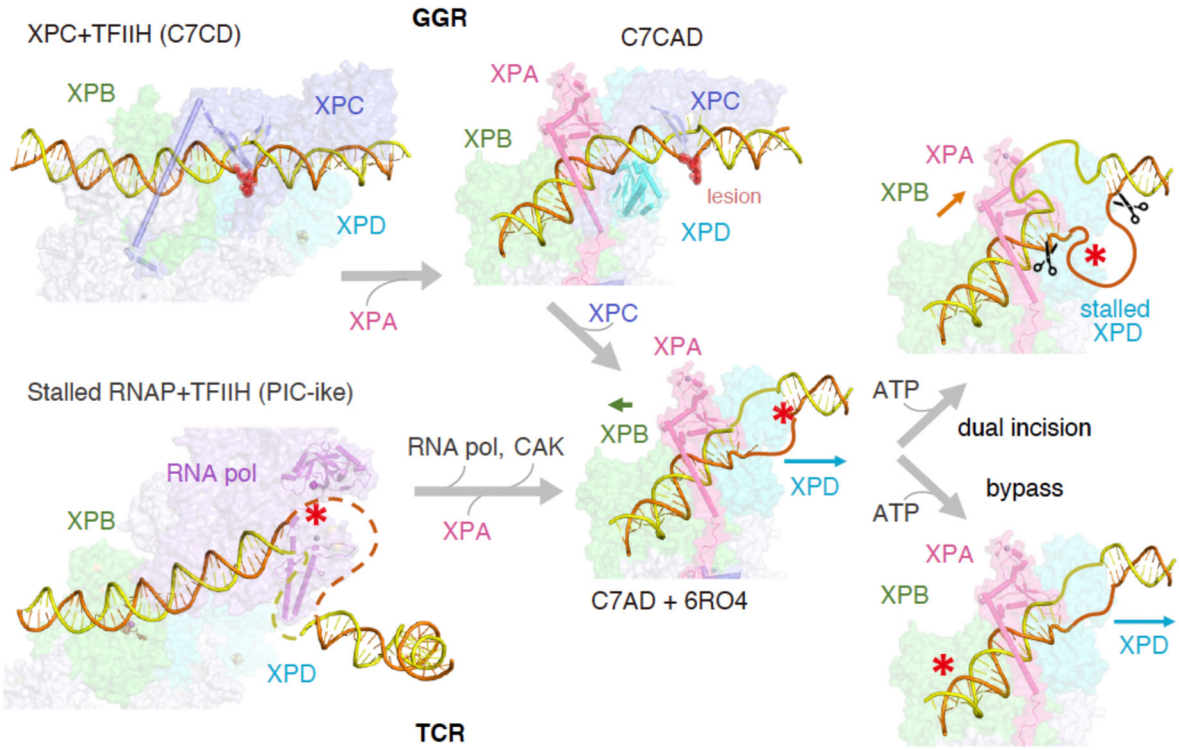


Fig. 5. Diagram of NER mechanism.

In GGR, lesion recognition by XPC and hand-off to TFIIH (Core7) and XPA are based on this study. In TCR, TFIIH recruitment is based on the PIC structure (PDB: 7NW0). In both pathways, a lesion is loaded 3′ to XPD. The ATPase activity of XPD leads TFIIH to translocate along ssDNA (indicated by the blue arrow) and scan for lesions (verification). A small lesion may be bypassed by XPD and allow continuous ssDNA scanning. A bulky DNA lesion would stall the XPD motor and render Core7 stationary. DNA upstream of the lesion is then translocated by XPB toward XPD (indicated by the orange arrow), and XPA separates the duplex and helps to enlarge the DNA bubble for dual incision.

Video 1.

Movement of XPC between flipping out 3 and 2 nt. The video was generated based on Cy5_C7CD and AP_C7CD (conf2). Protein subunits are color-coded according to Fig. 1a. The trimeric XPC translocates downstream of the lesion by 1 bp, which involves ~36° rotation and ~3.4 Å translation.

Video 2.

Movement of Core7 relative to XPC-DNA upon XPA binding. The video was generated based on Cy5_C7CD (C7CD) and Cy5_C7CAD. The attachment of the C-terminus of XPC to Core7 (XPB-p52-p8) prevents Core7 and XPC from dissociation. XPA is included in the last frame.

Video 3.

Structural changes between Cy5_C7CAD and AP_C7CAD. AP-DNA is more bent at the lesion site than Cy5, but it is more relaxed (straight) upstream of the lesion, where XPB and XPA bind.

Video 4.

Conformational changes from the U-shaped Core7 in C7CD to the σ shape in C7CAD. The XPB and XPD arms move relatively to the unchanged p52-p34-p44 interface at the U-turn.

Video 5.

An orthogonal view of the XPB and XPD movement between C7CD and C7CAD. This viewpoint shows the crossing of XPB and XPD subunits when forming the closed σ shape in C7CAD.

Video 6.

Movement of the XPD arm from C7CD to C7CAD. The flexible joint formed by p34-p44 in the XPD arm is stabilized by the p62 helical bundle.

Video 7.

Movement of the XPB arm from C7CD to C7CAD. The extensive domain interface between p52 and XPB allows XPB to undergo precession-like movement.