

ARTICLE

<https://doi.org/10.1038/s41467-019-10745-5>

OPEN

# Structural basis of TFIIH activation for nucleotide excision repair

Goran Kokic<sup>1</sup>, Aleksandar Chernev<sup>2,3</sup>, Dimitry Tegenov<sup>1</sup>, Christian Dienemann<sup>1</sup>, Henning Urlaub<sup>2,3</sup> & Patrick Cramer<sup>1</sup>

Nucleotide excision repair (NER) is the major DNA repair pathway that removes UV-induced and bulky DNA lesions. There is currently no structure of NER intermediates, which form around the large multisubunit transcription factor IIH (TFIIH). Here we report the cryo-EM structure of an NER intermediate containing TFIIH and the NER factor XPA. Compared to its transcription conformation, the TFIIH structure is rearranged such that its ATPase subunits XPB and XPD bind double- and single-stranded DNA, consistent with their translocase and helicase activities, respectively. XPA releases the inhibitory kinase module of TFIIH, displaces a 'plug' element from the DNA-binding pore in XPD, and together with the NER factor XPG stimulates XPD activity. Our results explain how TFIIH is switched from a transcription to a repair factor, and provide the basis for a mechanistic analysis of the NER pathway.

<sup>1</sup>Department of Molecular Biology, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany. <sup>2</sup>Max Planck Institute for Biophysical Chemistry, Bioanalytical Mass Spectrometry, Am Fassberg 11, 37077 Göttingen, Germany. <sup>3</sup>University Medical Center Göttingen, Institute of Clinical Chemistry, Bioanalytics Group, Robert-Koch-Straße 40, 37075 Göttingen, Germany. Correspondence and requests for materials should be addressed to P.C. (email: [patrick.cramer@mpibpc.mpg.de](mailto:patrick.cramer@mpibpc.mpg.de))

Genomes are constantly threatened by DNA damage, but cells can remove a large variety of DNA lesions by nucleotide excision repair (NER)<sup>1</sup>. Mutations in NER factors compromise cellular fitness and cause human diseases, such as Xeroderma pigmentosum (XP), Cockayne syndrome, and trichothiodystrophy (TTD)<sup>2,3</sup>. The NER machinery is built around the multisubunit transcription factor IIH (TFIIH), which opens the DNA repair bubble, scans for the lesion, and coordinates excision of the damaged DNA single-strand fragment<sup>1,4</sup>. TFIIH consists of a kinase module and a core module that contains the ATPases XPB and XPD<sup>5</sup>.

Here we prepare recombinant human TFIIH and show that XPB and XPD are stimulated by the additional NER factors XPA and XPG, respectively. We then determine the cryo-electron microscopy (cryo-EM) structure of the human core TFIIH-XPA-DNA complex at 3.6 Å resolution. The structure represents the lesion-scanning intermediate on the NER pathway and rationalizes the distinct phenotypes of disease mutations. It reveals that XPB and XPD bind double- and single-stranded DNA, respectively, and that XPA forms a bridge between XPB and XPD to retain the DNA at the 5'-edge of the repair bubble. Biochemical data and comparisons with prior structures<sup>6,7</sup> explain how XPA and XPG can switch TFIIH from a transcription factor to a DNA repair factor. During transcription, the kinase module inhibits the repair helicase XPD<sup>8</sup>. For DNA repair, XPA dramatically rearranges the core TFIIH structure, which reorients the ATPases, releases the kinase module, and displaces a “plug” element from the XPD pore that holds DNA. This enables XPD to move by ~80 Å, engage with DNA, and scan for the lesion.

## Results

**Recombinant human TFIIH.** A mechanistic dissection of the NER machinery was thus far hampered because TFIIH was not available in large quantities. We therefore established protocols to prepare milligram amounts of recombinant human TFIIH core and kinase modules (“Methods”). The purified TFIIH core comprised seven subunits including the ATPases XPB and XPD, whereas the kinase module contained CDK7, cyclin H, and MAT1. From these two modules, we could reconstitute the complete ten-subunit TFIIH (Supplementary Fig. 1a).

**Biochemical characterization of TFIIH ATPases.** To analyze the enzymatic activities of TFIIH, we monitored helicase and translocase activities in real time by fluorescence resonance energy transfer (“Methods”). Core TFIIH showed 5'-3' helicase activity, which was lost upon mutation of the XPD active site (Fig. 1a, Supplementary Fig. 1b). This is consistent with a prior description of XPD as a 5'-3' DNA helicase<sup>8,9</sup>. Core TFIIH also showed translocase activity, which was however much lower than the helicase activity (Fig. 1b). Translocase activity was due to XPB because it was sensitive to triptolide (Supplementary Fig. 1c), a drug that targets XPB<sup>10</sup>. This is consistent with the known translocase activity for the yeast XPB homolog<sup>11</sup> but not with helicase activity reported for an archaeal XPB homolog<sup>12</sup>.

**NER factors stimulate the TFIIH helicase XPD.** To test whether other NER factors affect TFIIH activities, we purified human XPA, XPG, RPA, and XPF-ERCC1 complex (Supplementary Fig. 1d). In the presence of XPA or XPG, DNA unwinding by XPD was 4-fold or 20-fold faster, respectively, as deduced from stopped-flow kinetics (Fig. 1a, Supplementary Fig. 1e). A stimulation of XPD by XPA was observed before<sup>9,13</sup>, but the effect of XPG on DNA unwinding is much stronger. This explains the earlier observation that XPG is required for efficient DNA bubble opening<sup>14</sup> and implicates XPG in lesion scanning by XPD. XPB

translocation activity was stimulated by all NER factors tested, although stimulation by XPA was exceptionally strong (Fig. 1b).

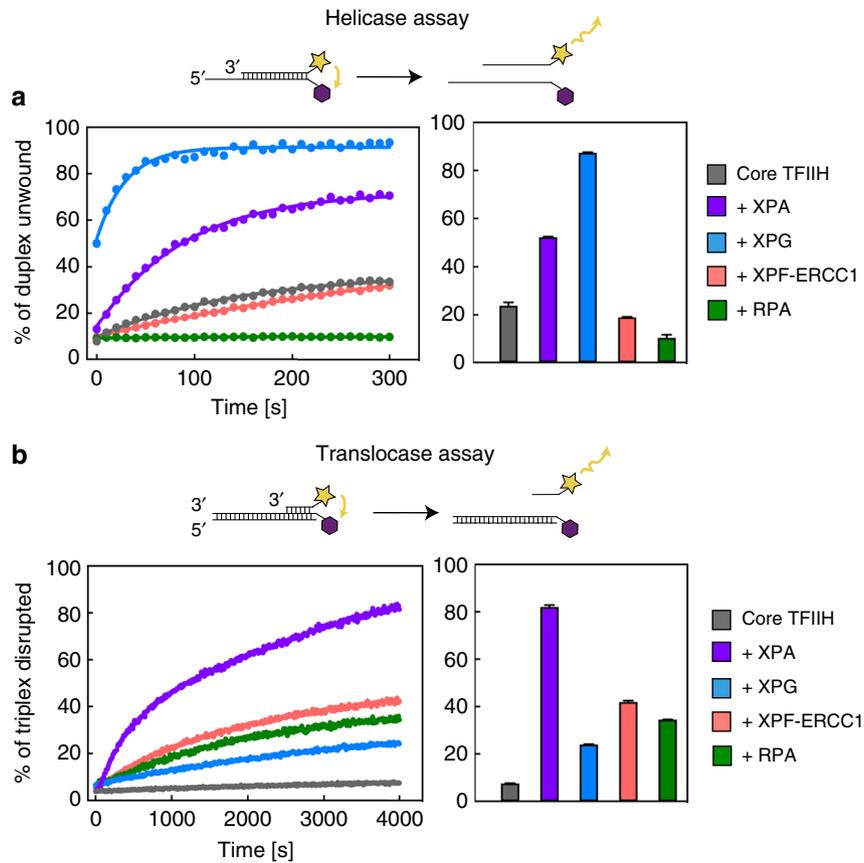
**Structure of the core TFIIH-XPA-DNA complex.** We next investigated the structural basis for how XPA and XPG activate the TFIIH ATPases. We prepared the core TFIIH-XPA-XPG complex bound to a bifurcated DNA scaffold that mimics one half of a DNA repair bubble (Supplementary Fig. 2a). We imaged this complex by cryo-EM and solved the structure at an overall resolution of 3.6 Å (“Methods,” Supplementary Figs. 2 and 3). The cryo-EM density was of high quality and revealed DNA and all protein components except XPG, which likely dissociated during cryo-EM grid preparation. The derived structure contains the p52 subunit and other regions that were lacking from the previous human TFIIH structure<sup>6</sup> and reveals the XPB-TFIIH core interface (Fig. 2, Supplementary Fig. 3f).

**TFIIH rearrangements.** The structure of the core TFIIH-XPA-DNA complex differs substantially from the TFIIH structure observed in transcription complexes (Fig. 2)<sup>7,15</sup>. Both XPB and XPD bind DNA (Supplementary Fig. 4), whereas only XPB binds DNA in transcription complexes<sup>7,15</sup>. XPB binds DNA in the duplex region, whereas XPD binds the single-stranded 3'-DNA extension, consistent with translocase and helicase function, respectively. DNA binding of both ATPases requires large structural changes in TFIIH (Supplementary Movie 1). XPD and its associated subunit p44 move by ~80 Å, and this requires a flexible connection between subunits p44 and p34 and rearrangements in subunit p52 (Supplementary Fig. 5).

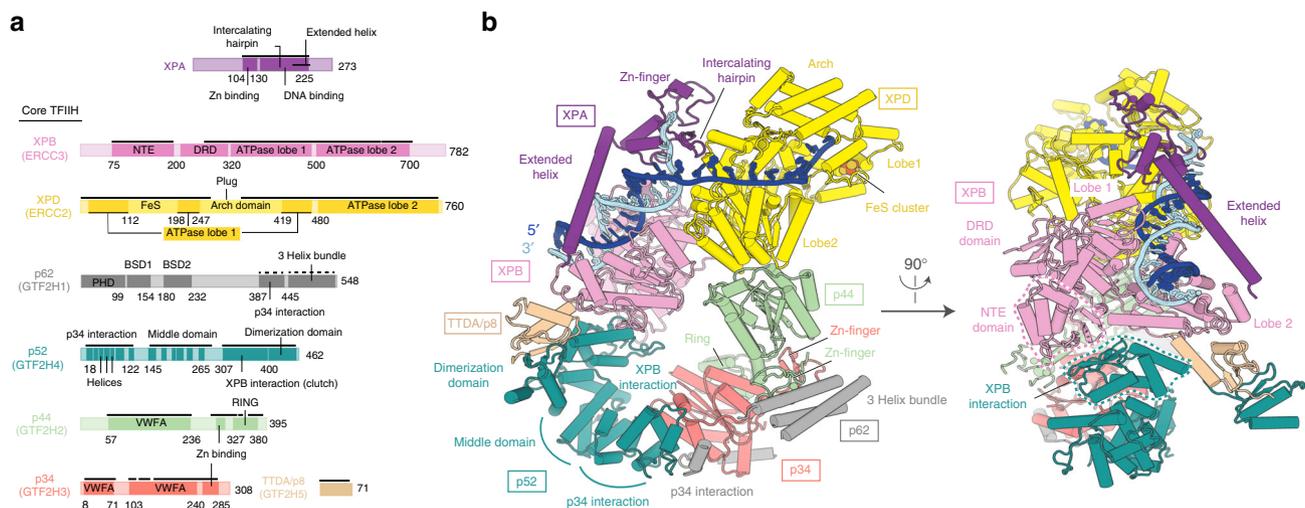
**XPA clamps TFIIH to DNA.** The structure informs on XPA, which is essential for NER<sup>16</sup>. XPA contains an N-terminal zinc finger and a DNA-binding domain with an extended helix and an intercalating  $\beta$ -hairpin (Fig. 2). XPA forms an elongated arch over DNA that bridges between the two ATPases. XPA binds XPB and XPD with its extended helix and its intercalating hairpin, respectively. The C-terminal region of XPA extends to p52 and TTDA/p8 (Supplementary Figs. 3d and 7c), explaining why TTDA/p8 facilitates XPA recruitment *in vivo*<sup>17</sup>.

These observations explain how XPA stimulates XPB translocation. First, XPA connects both XPB ATPase lobes to p52 and TTDA/p8 subunits that stimulate the XPB activity within the TFIIH core<sup>18,19</sup>. Second, XPB is not a processive enzyme and readily dissociates from DNA<sup>11</sup>. However, in our structure DNA is held in a positively charged DNA duplex tunnel that is formed between the extended helix of XPA and the XPB ATPase (Fig. 3a). XPA thus retains DNA near the XPB active site. Indeed, our cryo-EM data revealed an alternative state of the complex with the DNA duplex disengaged from XPB but retained by the XPA extended helix (Fig. 3b, Supplementary Fig. 2d). Thus, by trapping the DNA within the duplex tunnel, XPA may retain the NER machinery on the DNA during lesion scanning and processing.

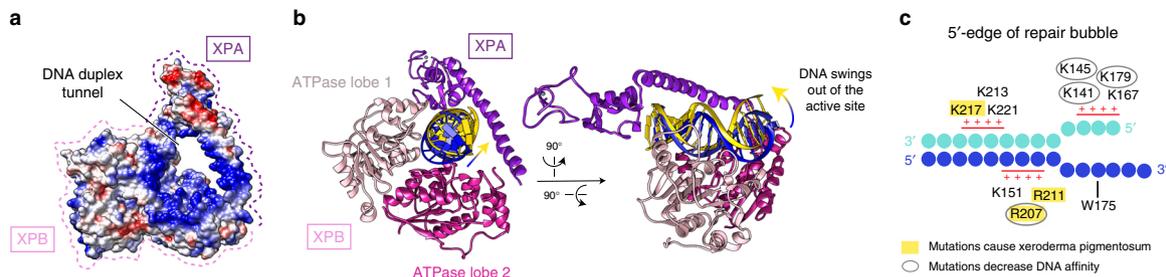
**XPA demarcates the edge of the DNA repair bubble.** XPA also contributes to the recognition of the 5'-edge of the DNA repair bubble that depends on electrostatic interactions (Fig. 3c). XPA inserts its intercalating  $\beta$ -hairpin between DNA single strands at the duplex junction (Fig. 2b), consistent with the published biochemical data<sup>20</sup>. The tip of the XPA hairpin contains a conserved tryptophan residue (Trp175) that stacks against the base of the DNA 3'-extension at the junction (Supplementary Fig. 3i). Several sites of mutations that cause severe XP<sup>16</sup> map to XPA residues that interact with DNA (Fig. 3c). Previous studies of the yeast XPA counterpart suggested that XPA may detect DNA lesions<sup>21</sup>.



**Fig. 1** Regulation of ATPases in core transcription factor IIH. **a** Effect of nucleotide excision repair (NER) proteins on XPD 5'-3' helicase activity. Real-time fluorescence measurement using a fluorescence energy transfer-based assay. Bars show the percentage of unwound product after 100 s ( $n = 2$ , error bars indicate the range of the data). RPA inhibits XPD helicase activity, by masking the single-stranded DNA overhang. Source Data are provided in the Source Data file. **b** Effect of NER proteins on XPB translocase activity. Real-time fluorescence measurement of triplex disruption. Bars show percentage of disrupted triplex after 4000 s ( $n = 2$ , error bars indicate the range of the data). Source Data are provided in the Source Data file



**Fig. 2** Structure of human core transcription factor IIH (TFIIH)-XPA-DNA complex. **a** Domain organization of XPA and human TFIIH subunits. Residues at domain borders are indicated. Solid and dashed black lines mark residues modeled as atomic and backbone structures, respectively. DRD damage recognition domain, NTE N-terminal extension, BSD BTF2-like, synapse-associated and DOS2-like domains, VWFA von Willebrand factor type A domain. **b** Cylindrical representation of the structure. Proteins colored as in **a**. Newly built XPB and p52 domains are highlighted with dashed lines and reveal how XPB is connected to the TFIIH core



**Fig. 3** XPA-DNA interactions. **a** DNA duplex tunnel formed by XPA and XPB. Blue, white, and red color indicates positive, neutral, and negative electrostatic surface potential, respectively. Created with UCSF Chimera. **b** Two positions of DNA in the tunnel. Tightly bound DNA is in blue, dissociated DNA in yellow, ATPase lobe 1 of XPB in pink, ATPase lobe 2 in hot pink, and XPA in purple. **c** Electrostatic interactions between XPA and the DNA junction. DNA nucleotides are indicated as circles. Patches of positively charged residues in proximity to the DNA backbone are indicated. Residues that are mutated in Xeroderma pigmentosum are highlighted in yellow<sup>16</sup>. Mutation of encircled residues decreases DNA affinity<sup>61</sup>

However, our results suggest that XPA rather demarcates the 5'-edge of the repair bubble and stimulates lesion scanning by clamping core TFIIH onto DNA.

**Disease-related XPD mutations.** The detailed structure also enables the localization of XPD residues mutated in patients with XP and TTD and rationalizes their functional effects, as previously suggested by biochemical studies<sup>22</sup> and comparisons to archaeal XPD homologs in the absence of DNA<sup>23–25</sup> (Supplementary Fig. 6, Supplementary Note 1). Most notably, the XP mutations affect the DNA-contacting residues (Supplementary Fig. 6b) in the ATPase lobe 2, which would specifically impair the XPD helicase activity and DNA repair. In contrast, TTD mutations map to the XPD-p44 interface, around the FeS cluster and between the XPD ATPase lobes (Supplementary Fig. 6c–e). Thus TTD mutations would compromise the integrity of TFIIH and XPD stability, thereby affecting transcription and other TFIIH-mediated processes outside NER<sup>5</sup>, which leads to more severe phenotypes.

**XPD-DNA interactions.** The structure also provides details of the XPD-DNA interactions (Fig. 4a, b). The ATPase lobe 2 interacts with DNA bases near the duplex-single-strand junction, which includes base stacking with the side chains of residues F508 and Y627. This mode of DNA interaction is unusual for helicases of the SF2 family, which generally engage with the sugar-phosphate backbone<sup>26,27</sup>. We speculate that the extensive contacts of XPA and XPD with single-stranded DNA facilitate DNA opening and XPD loading during initial stages of NER<sup>28</sup>.

The structure further suggests how XPD verifies the lesion during DNA scanning. The DNA single strand extends into a pore formed by the ATPase lobe 1, the iron-sulfur cluster (FeS) domain, and the Arch domain (Fig. 4a). The sugar-phosphate backbone is bound by residues in the FeS domain, including Y192 and R196, which were implicated in DNA lesion sensing<sup>29</sup> (Fig. 4b). Residues R112 and C134 bridge between DNA and the FeS cluster (Supplementary Fig. 4d), which was suggested to be involved in lesion detection via DNA-mediated charge transfer<sup>30</sup>. The FeS cluster is flanked by two protein pockets that are lined with aromatic residues Y158, F161, and F193 and may proof-read DNA bases, as observed for base excision DNA repair<sup>31</sup>.

**Mechanism of XPD repression.** We could reproduce the known repression of the XPD helicase activity by the TFIIH kinase module<sup>8,9</sup> in our helicase assays (Fig. 4d). We found that catalytically inactive variants of the TFIIH kinase module (Supplementary Fig. 1f) could also repress XPD helicase activity, showing that repression does not require the kinase activity of CDK7

(Fig. 4d). Both XPA and XPG could relieve the kinase-mediated repression of XPD in a concentration-dependent manner (Fig. 4e). These observations show that XPA and XPG counteract the repressive effect of the kinase module on XPD.

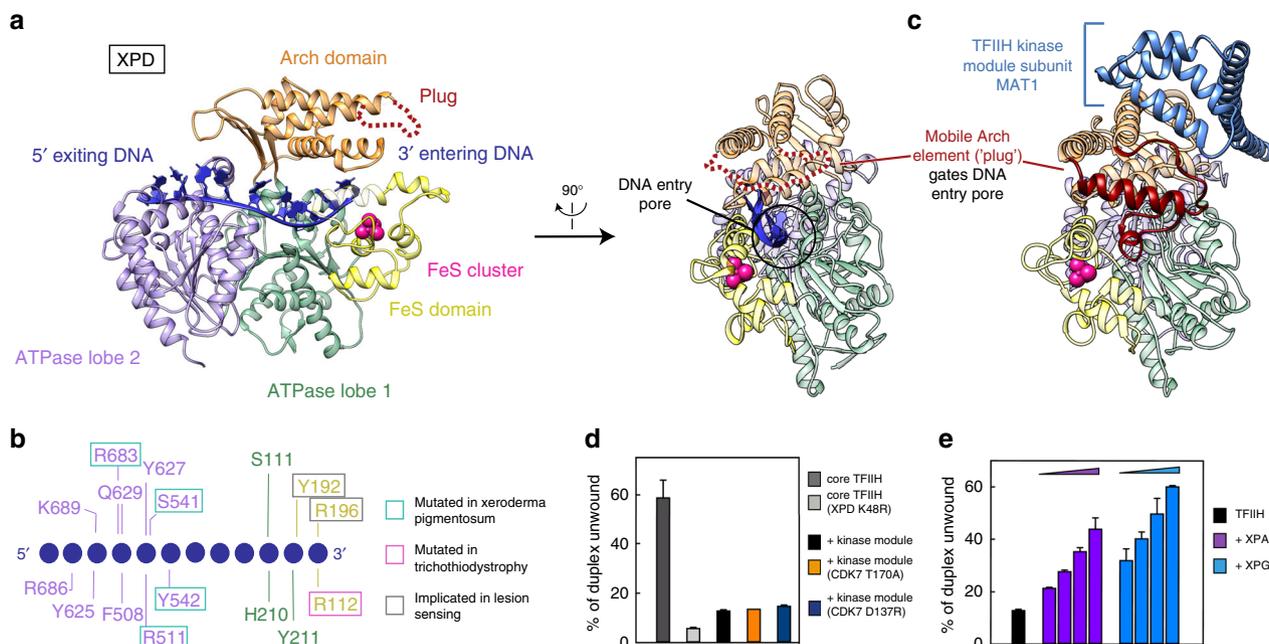
A comparison of our structure with the previous TFIIH structures shows how the kinase module represses XPD activity. In previous TFIIH structures<sup>6,7</sup>, a region in the XPD Arch domain forms a “plug” (residues 273–325) that occupies the DNA pore of XPD (Fig. 4c). The plug would clash with DNA in the XPD pore but is displaced and mobile in our structure (Fig. 4a). The kinase module subunit MAT1 contacts the plug and may stabilize it in the XPD pore, explaining how the kinase module impairs binding of core TFIIH to single-stranded DNA and XPD helicase activity<sup>9</sup>. In addition, a loop in the yeast counterpart of p62 extends into the XPD active site<sup>7</sup> and would interfere with the observed DNA trajectory through the helicase.

**Mechanism of XPD activation by XPA.** Structural comparisons also suggest how XPA relieves XPD inhibition by the kinase module. XPA stabilizes TFIIH in a new conformation in which the two ATPases are drastically reoriented. This conformation is incompatible with MAT1 binding as observed in the previous TFIIH structure<sup>6</sup> (Supplementary Fig. 5e). This also explains how XPA facilitates kinase module removal upon NER induction *in vivo*<sup>32</sup>. Taken together, MAT1 and XPA stabilize two entirely different conformations of TFIIH, which contain the repair helicase XPD in an inactive or an active state, respectively.

**XPG in lesion scanning.** Since XPG was not visible in our structure, we located it by chemical crosslinking (Supplementary Fig. 7, Supplementary Data 1). The crosslinking data match the structural data very well (Supplementary Fig. 7c), and unambiguously localize XPG. The N-terminal region of XPG specifically crosslinks to the FeS and Arch domains in XPD, including the plug element, whereas the C-terminal extension of XPG crosslinks mostly to XPB and p52 (Supplementary Fig. 7a, d). In addition, XPG crosslinks to XPD at its binding site for the kinase module (Supplementary Fig. 7d), suggesting that XPG competes with the kinase module for XPD binding and explaining how XPD inhibition by the kinase module is relieved (Fig. 4e). These data suggest that XPG facilitates lesion scanning by blocking the kinase module-binding site on XPD and directly stimulating XPD helicase activity. We note that XPG may bind to TFIIH in alternative ways during other TFIIH functions, in which XPD is not bound to DNA<sup>5–7,33</sup>.

## Discussion

Taken together, our structure-function analysis extends our understanding of the NER pathway (Supplementary Fig. 8). The



**Fig. 4** XPD activation and DNA binding. **a** Two views of XPD bound to DNA. XPD domains ATPase lobe 1, FeS, Arch, and lobe 2 are in green, yellow, orange, and medium purple, respectively. DNA is dark blue. A black circle depicts the DNA pore. **b** Schematic representation of XPD-DNA interactions. **c** Side view of XPD bound to the kinase module (PDB code 5OF4)<sup>6</sup>. The plug in the Arch domain is in dark red, the kinase module subunit MAT1 in blue. **d** Effect of kinase module variants on XPD helicase activity. Core transcription factor IIH (TFIIH) was incubated with two-fold excess of the kinase module and helicase activity was monitored as in Fig. 1a. Bars show the percentage of unwound product after 300 s ( $n = 2$ , error bars indicate the range of the data). Source Data are provided in the Source Data file. **e** Effect of increasing concentrations of XPA and XPG on XPD helicase activity in the presence of kinase module (0.25  $\mu\text{M}$  core TFIIH, 0.5  $\mu\text{M}$  kinase module, 0.375, 0.75, 1.5, or 3  $\mu\text{M}$  XPA or XPG). Otherwise as in **d**. Source Data are provided in the Source Data file

DNA lesion is first recognized by XPC, which recruits TFIIH<sup>1,13</sup>. XPA and XPG then displace the kinase module, stabilize an alternative conformation of TFIIH, and remove the inhibitory plug from the XPD pore. XPA and XPG also stimulate XPB and XPD, and this may facilitate DNA opening<sup>4,14,18,19</sup> and XPD migration in the 3' direction to scan the DNA strand for the lesion<sup>13</sup>. XPA chaperons TFIIH-DNA interactions and anchors the NER machinery to the 5'-edge of the repair bubble, where it is ideally positioned to recruit XPF-ERCC1 and complete the repair assembly when the lesion is encountered<sup>16</sup>. The two endonucleases XPF-ERCC1 and XPG can now incise the lesion-containing DNA strand near the 5'- and the 3'-edge of the repair bubble, respectively, to remove the lesion-containing DNA fragment<sup>1</sup>. We acknowledge that, while our manuscript was in revision, a manuscript became available online<sup>34</sup> that provides a high-resolution free TFIIH structure and includes a mapping of disease mutations consistent with the one described here.

## Methods

**Cloning and protein expression.** Vectors encoding full-length XPA, XPG, XPF, ERCC1, RPA1, RPA2, RPA3, XPB, XPD, p62, p52, p44, p34, MAT1, CDK7, cyclin H, and TTDA were ordered from Harvard Medical School Plasmid Repository and used as a template for gene amplification by PCR. Amplified genes were cloned into MacroBac vectors via ligation-independent cloning, as described<sup>35</sup>. ERCC1, p52, p34, and TTDA were cloned into 438A (Addgene: 55218); XPA, XPF, XPB, p62, p44, MAT1, CDK7, and Cych into 438B (Addgene: 55219); and XPG and XPD into 438C vector (Addgene: 55220), which resulted in no tag, N-terminal 6 $\times$  His or N-terminal 6 $\times$  His followed by a maltose-binding protein, respectively. XPF and ERCC1, MAT1, CDK7, and Cych, as well as XPB, p62, p52, p44, p34, and TTDA were combined into a single vector via restriction digestion and ligation-independent cloning<sup>35</sup>. RPA1 was cloned in 11B (Addgene: 48295), and RPA 2 and 3 were cloned into 11A vectors (Addgene: 48294), followed by assembly of all RPA subunits into one vector by successive ligation-independent cloning reactions. All tags were separated from the gene with a tobacco etch virus protease site. All mutant proteins used here (CDK7:D137R, CDK7:T170A, XPG:E791A, XPD:K48R)

were produced by round-the-horn side-directed mutagenesis and purified as their wild-type counterparts.

RPA complex was expressed in *Escherichia coli* and purified as described<sup>36</sup>. TFIIH kinase module comprised of MAT1, CDK7, or CDK7 mutants and cyclin H was expressed and purified as described<sup>37</sup>. All other proteins and protein complexes were expressed in insect cells and purified as described below. Bacmid preparation and V0 and V1 virus production was as described previously<sup>38</sup>. Core TFIIH was produced by co-infecting the cells with two V1 viruses: virus encoding XPB, p62, p52, p44, p34, and TTDA and a virus encoding XPD or XPD mutant. All proteins were expressed in Hi5 cells grown in ESF-921 media (Expression Systems, Davis, CA, USA) at 27  $^{\circ}\text{C}$ . Typically, 600 ml culture was infected with 500  $\mu\text{l}$  of V1 virus and grown for 48–72 h prior harvesting by centrifugation (30 min, 4  $^{\circ}\text{C}$ , 500  $\times$  g). Cell pellet was resuspended in lysis buffer; 400 mM KCl, 20% glycerol (v/v), 20 mM KOH-HEPES pH 7.0, 5 mM  $\beta$ -mercaptoethanol, 30 mM imidazole pH 8.0, 0.284  $\mu\text{g ml}^{-1}$  leupeptin, 1.37  $\mu\text{g ml}^{-1}$  pepstatin A, 0.17 mg  $\text{ml}^{-1}$  phenylmethanesulfonylfluoride (PMSF), and 0.33 mg  $\text{ml}^{-1}$  benzimidazole for the core TFIIH and 400 mM NaCl, 20 mM Tris-HCl pH 7.9, 10% glycerol (v/v), 1 mM dithiothreitol (DTT), 30 mM imidazole pH 8.0, 0.284  $\mu\text{g ml}^{-1}$  leupeptin, 1.37  $\mu\text{g ml}^{-1}$  pepstatin A, 0.17 mg  $\text{ml}^{-1}$  PMSF, and 0.33 mg  $\text{ml}^{-1}$  benzimidazole for the other proteins. Cell suspension was flash frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$ .

**Protein purification.** All purification steps were performed at 4  $^{\circ}\text{C}$  and all buffers were filtered and thoroughly degassed immediately before use. Cells were thawed in a water bath operating at 30  $^{\circ}\text{C}$  and opened by sonication. The lysate was clarified by centrifugation (18,000  $\times$  g, 30 min), followed by ultracentrifugation (235,000  $\times$  g, 60 min). In case of core TFIIH, the clarified lysate was first filtered using 0.8- $\mu\text{m}$  syringe filters (Millipore) and loaded onto HisTrap HP 5 ml (GE Healthcare, Little Chalfont, UK). The column was washed with 10 column volumes (CV) of lysis buffer, followed by 20 CV of high salt wash (800 mM KCl, 20% glycerol (v/v), 20 mM KOH-HEPES pH 7.0, 5 mM  $\beta$ -mercaptoethanol, 30 mM imidazole pH 8.0, 0.284  $\mu\text{g ml}^{-1}$  leupeptin, 1.37  $\mu\text{g ml}^{-1}$  pepstatin A, 0.17 mg  $\text{ml}^{-1}$  PMSF, and 0.33 mg  $\text{ml}^{-1}$  benzimidazole). Column was washed again with 5 CV of lysis buffer and protein was subsequently eluted with a gradient of 0–80% elution buffer (400 mM KCl, 20% glycerol (v/v), 20 mM KOH-HEPES pH 7.0, 5 mM  $\beta$ -mercaptoethanol, 500 mM imidazole pH 8.0, 0.284  $\mu\text{g ml}^{-1}$  leupeptin, 1.37  $\mu\text{g ml}^{-1}$  pepstatin A, 0.17 mg  $\text{ml}^{-1}$  PMSF, and 0.33 mg  $\text{ml}^{-1}$  benzimidazole). Fractions were checked on NuPAGE 4–12% Bis-Tris Protein Gels (Invitrogen) for the presence of all core TFIIH subunits, and appropriate fractions were pulled and mixed with 10 ml of amylose resin (New England BioLabs) pre-equilibrated in washing buffer (400 mM KCl, 20% glycerol (v/v), 20 mM KOH-

HEPES pH 7.0, 5 mM  $\beta$ -mercaptoethanol, 2 mM  $MgCl_2$ , and 10  $\mu M$   $ZnCl_2$ ). The protein solution was incubated with the beads for 1 h while rotating. The amylose resin was poured into Econo-Pac Chromatography columns (BioRad) and washed with 5 CV of washing buffer. Protein was eluted with washing buffer containing 100 mM maltose. Protein-containing fractions were pooled, mixed with 2 mg of TEV protease, and dialyzed against 2 l of dialysis buffer overnight (250 mM KCl, 20% glycerol (v/v), 20 mM KOH-HEPES pH 7.0, 5 mM  $\beta$ -mercaptoethanol, 2 mM  $MgCl_2$ , and 10  $\mu M$   $ZnCl_2$ ). The dialyzed sample was applied to DEAE (GE Healthcare) and heparin column (GE Healthcare) in tandem and washed with 20 CV of dialysis buffer. After the removal of DEAE column, protein was eluted with a gradient of elution buffer 0–100% (1 M KCl, 20% glycerol (v/v), 20 mM KOH-HEPES pH 7.0, 5 mM  $\beta$ -mercaptoethanol, 2 mM  $MgCl_2$ , and 10  $\mu M$   $ZnCl_2$ ). Peak fractions were pooled, concentrated with Amicon Millipore 15 ml 100,000 MWCO centrifugal concentrator, and applied to Superose 6 increase 10/300 GL column (GE Healthcare) equilibrated in storage buffer (400 mM KCl, 20% glycerol (v/v), 20 mM KOH-HEPES pH 7.0, 5 mM  $\beta$ -mercaptoethanol, 2 mM  $MgCl_2$ ). Peak fractions were again concentrated, aliquoted, flash frozen, and stored at  $-80^\circ C$ .

XPF-ERCC1-, XPG-, and XPA-containing lysate was applied onto GE HisTrap HP 5 ml (GE Healthcare, Little Chalfont, UK) equilibrated in lysis buffer (in case of XPA, all downstream steps were performed in the presence of 5 mM  $\beta$ -mercaptoethanol and 10  $\mu M$   $ZnCl_2$  instead of 1 mM DTT). The column was washed with 20 CV of high salt buffer (800 mM NaCl, 20 mM Tris-HCl pH 7.9, 10% glycerol (v/v), 1 mM DTT, 30 mM imidazole pH 8.0, 0.284  $\mu g$  ml $^{-1}$  leupeptin, 1.37  $\mu g$  ml $^{-1}$  pepstatin A, 0.17 mg ml $^{-1}$  PMSF, and 0.33 mg ml $^{-1}$  benzamide). After 5 CV wash with the lysis buffer, proteins were eluted with the elution buffer gradient 0–80% (400 mM NaCl, 20 mM Tris-HCl pH 7.9, 10% glycerol (v/v), 1 mM DTT, 500 mM imidazole pH 8.0, 0.284  $\mu g$  ml $^{-1}$  leupeptin, 1.37  $\mu g$  ml $^{-1}$  pepstatin A, 0.17 mg ml $^{-1}$  PMSF, and 0.33 mg ml $^{-1}$  benzamide). Pulled peak fractions were processed differently for different proteins. XPF-ERCC1- and XPA-containing protein solutions were directly mixed with 2 mg of TEV protease and dialyzed against 1 liter of dialysis buffer (400 mM NaCl, 20 mM Tris-HCl pH 7.9, 10% glycerol (v/v), 1 mM DTT). XPG solution was mixed with 10 ml of amylose resin (New England BioLabs) pre-equilibrated in dialysis buffer. Solution was mixed for 1 h and subsequently poured into Econo-Pac Chromatography columns (BioRad). Column was washed with 10 CV of dialysis buffer followed by elution with dialysis buffer containing 100 mM maltose. Protein-containing fractions were pulled, mixed with 2 mg of TEV protease, and dialyzed against 1 liter of dialysis buffer. Dialyzed solutions containing XPF-ERCC1, XPA, and XPG were loaded on GE HisTrap HP 5 ml and flow through fractions were collected. Protein-containing fractions were checked for contaminants on NuPAGE 4–12% Bis-Tris Protein Gels (Invitrogen), pulled, concentrated with appropriate Amicon Millipore 15 ml centrifugal concentrator, and applied onto Superdex 75 10/300 equilibrated in storage buffer (400 mM NaCl, 20 mM NaOH:HEPES pH 7.5, 10% glycerol (v/v), 10  $\mu M$   $ZnCl_2$ , and 5 mM  $\beta$ -mercaptoethanol) for XPA and Superdex 200 10/300 increase (GE Healthcare) equilibrated in a different storage buffer (400 mM NaCl, 20 mM NaOH:HEPES pH 7.5, 10% glycerol (v/v), 1 mM DTT) for the rest of the proteins. Peak fractions were pooled, concentrated, aliquoted, flash frozen, and stored at  $-80^\circ C$ .

**Helicase and translocase assays.** H1 and H2 DNA sequences (Supplementary Table 1) were used for monitoring the helicase activity in 5'–3' direction and H3 and H4 for monitoring the helicase activity in 3'–5' direction. DNA annealing reaction contained fluorescent DNA primer (25  $\mu M$ ) and quenching DNA oligo (37.5  $\mu M$ ) dissolved in water. Annealing was performed in a thermocycler by heating up the DNA solution to 95  $^\circ C$  for 5 min, followed by slow cooling (1  $^\circ C$ /min) to 4  $^\circ C$ . Typical unwinding reactions of 20  $\mu l$  final volume contained 0.4 pmol of DNA duplex and 8 pmol of core TFIID in 100 mM KCl, 20 mM KOH:HEPES pH 7.0, 5% glycerol, 0.2 mg ml $^{-1}$  bovine serum albumin, 3 mM phosphoenolpyruvate, 10 mM  $MgCl_2$ , 1 mM DTT, and excess amount of pyruvate kinase (Sigma). When the effect of DNA repair factors on unwinding was measured, we supplemented the reaction with 24 pmol of the corresponding factor. The reaction mixture was preincubated at 26  $^\circ C$  for 10 min. The reaction was started by addition of ATP (2 mM final) and the unwinding was monitored at 26  $^\circ C$  by using the Infinite M1000Pro reader with excitation wavelength 495 nm, emission wavelength 520 nm, and gain of 150. Percentage of unwound product was calculated by dividing the observed fluorescence intensity by the intensity of the fluorescent primer in the reaction buffer (mimicking fully unwound DNA).

DNA unwinding monitored by stopped-flow was performed in the same buffer conditions and with the same final protein and DNA concentrations as above. The core TFIID preincubated with XPA or XPG was rapidly mixed with equal volume of ATP (2 mM final) in the SX-20MV stopped-flow apparatus (Applied Photophysics). FAM fluorescence was monitored upon excitation at 465 nm after passing through KV500 cut-off filter (Schott). All time courses shown represent average of five technical replicates. Initial rate of DNA unwinding was calculated using Prism (Graphpad software) by fitting the initial linear part of the fluorescence trace.

Triplex displacement assay was performed in a similar way as previously described<sup>11</sup>. Ten-microliter annealing reaction for triplex displacement assay contained T1 (30  $\mu M$ ) and T2 (25  $\mu M$ ) oligo (Supplementary Table 1) in 25 mM

MES pH 5.5 and 10 mM  $MgCl_2$ . The reaction was heated to 95  $^\circ C$  for 5 min followed by slow cooling (1  $^\circ C$ /min) to 4  $^\circ C$ . After cooling, the reaction was supplemented with 1  $\mu l$  of fluorescent T3 oligo (9  $\mu M$  final), heated to 57  $^\circ C$ , and cooled down to 20  $^\circ C$  at the speed of 1  $^\circ C$ /min. Translocation reactions were performed exactly as described for the helicase assay, only with triplex DNA as a substrate. A higher core TFIID input (75 pmol in 20  $\mu l$  reactions) was used when no stimulatory factors were added (Supplementary Fig. 1c) to obtain a more robust fluorescent signal.

**Kinase activity assay.** We used the kinase activity assay to assess the activity of kinase module variants containing CDK7:D137R or CDK7:T170A mutants<sup>39</sup>. As CDK7 phosphorylates the C-terminal domain of the largest RNA polymerase II subunit during transcription initiation, we used purified yeast RNA polymerase II<sup>40</sup> dephosphorylated with lambda phosphatase during purification<sup>37</sup> as a substrate in the assay. RNA polymerase II (50 nM final) was mixed with increasing concentrations of kinase module variants (30, 100, 220, and 500 nM final) in final buffer conditions containing 100 mM KCl, 20 mM KOH:HEPES pH 7.5, 3 mM  $MgCl_2$ , 5% glycerol, and 5 mM  $\beta$ -mercaptoethanol and preincubated for 2 min at 30  $^\circ C$ . Reactions were started by the addition of ATP (0.5 mM final) and quenched after 2 min at 30  $^\circ C$  with EDTA (100 mM) and 4 $\times$  LDS buffer (Invitrogen). Reactions were run on 4–12% Bis-Tris gel in MOPS buffer (ThermoFisher Scientific) and transferred to nitrocellulose membranes (GE Healthcare Life Sciences). The membranes were blocked with 5% (w/v) milk in phosphate-buffered saline (PBS) buffer supplemented with 0.1% Tween 20 for 1 h at room temperature. The membranes were treated with primary antibody (3E8, 1:25 dilution) in 0.25% (w/v) milk in PBS supplemented with 0.1% Tween 20 and incubated at room temperature for 1 h. After several rounds of washing with PBS buffer supplemented with 0.1% Tween 20, the membranes were incubated with horseradish peroxidase-conjugated anti-rat secondary antibody (1:5000 dilution, Sigma-Aldrich A9037) in 0.1% (w/v) milk in PBS supplemented with 0.1% Tween 20 and incubated at room temperature for 1 h. Antibodies were detected with SuperSignal West Pico Chemiluminescent Substrate (ThermoFisher) and the membranes were scanned with ChemoCam Advanced Fluorescence imaging system (Intas Science Imaging).

**Mass spectrometric identification of crosslinking sites.** Sample for crosslinking was prepared by mixing core TFIID, XPA, XPG, and bifurcated DNA scaffold (see “Cryo-EM sample preparation and image processing”) in 1:2:2:1.5 molar ratio in a final buffer containing 150 mM KCl, 10% glycerol, 2 mM  $MgCl_2$ , 20 mM KOH:HEPES pH 7.5, and 5 mM  $\beta$ -mercaptoethanol. The reaction was incubated for 20 min at room temperature before applying to the Superose 6 increase 3.2/300 column equilibrated in final buffer used for the complex formation. Fractions were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis and Coomassie staining. For crosslinking, purified complexes (~2.2  $\mu M$  final) were supplemented with BS3 crosslinker (2 mM final, ThermoFisher Scientific) and incubated for 30 min at 30  $^\circ C$ . Reaction was quenched with ammonium bicarbonate (200 mM final) and further incubated for 10 min at 30  $^\circ C$ .

Crosslinked proteins were reduced with 10 mM DTT for 30 min at 37  $^\circ C$ , followed by alkylation with 40 mM iodoacetamide for 30 min at 25  $^\circ C$ . The proteins were digested overnight at 37  $^\circ C$  in the presence of 1 M urea with trypsin in 1:20 (w/w) enzyme-to-protein ratio. The samples were acidified with formic acid (FA) to 0.1% (v/v) final concentration and acetonitrile (ACN) was added to 5% (v/v) final concentration. Peptides were bound to Sep-Pak C18 50 mg sorbent cartridge (Waters), washed with 5% ACN and 0.1% FA (v/v), eluted with 80% ACN and 0.1% FA (v/v), dried under vacuum, and resuspended in 30  $\mu l$  30% ACN and 0.1% trifluoroacetic acid (TFA) (v/v). The sample was fractionated on Superdex Peptide PC3.2/30 column (GE Healthcare) at a flow rate of 50  $\mu l$  min $^{-1}$  of 30% ACN, 0.1% TFA (v/v) and 100  $\mu l$  fractions corresponding to elution volume 0.9–2.2 ml were collected, dried under vacuum, and resuspended in 20  $\mu l$  2% ACN and 0.05% TFA (v/v).

Crosslinked peptides were analyzed on Orbitrap Fusion Tribrid Mass Spectrometer (Thermo Scientific), coupled to Dionex UltiMate 3000 UHPLC system (Thermo Scientific) equipped with an in-house-packed C18 column (ReproSil-Pur 120 C18-AQ, 1.9  $\mu m$  pore size, 75  $\mu m$  inner diameter, 30 cm length, Dr. Maisch GmbH). Samples were analyzed as 5  $\mu l$  injections, separated on 118 min gradient: mobile phase A—0.1% FA (v/v); mobile phase B—80% ACN, 0.08% FA (v/v). The gradient started at 8% B and increased to 48% B in 106 min, then keeping B constant at 90% for 6 min, followed by re-equilibration of the column with 5% B. The flow rate was set to 300 nl min $^{-1}$ . MS1 spectra were acquired with the following settings: resolution—120,000; mass analyzer—Orbitrap; mass range—380–1500  $m/z$ ; injection time—60 ms; automatic gain control target— $4 \times 10^5$ . Dynamic exclusion was set to 9 s. For samples of elution volume up to 1.7 ml, only charges 3–8 were included. For subsequent samples, charges 2–8 were included. MS2 spectra settings: resolution—30,000; mass analyzer—Orbitrap; injection time—128 ms; automatic gain control target— $5 \times 10^4$ ; isolation window—1.6  $m/z$ . Fragmentation was enforced by higher-energy collisional dissociation and varied for the three injections at 30%, 28%, and stepped collision energy with step +/- 2% at 28%.

Raw files were converted to mgf format with ProteomeDiscoverer 2.1.0.81 (Thermo Scientific); signal-to-noise ratio—1.5; precursor mass—350–7000 Da. For

identification of crosslinked peptides, files were analyzed by pLink (v. 1.23) (pFind group)<sup>41</sup> with the following settings: crosslinker—BS3 (default settings—lysine and the protein N-terminus); digestion enzyme—trypsin; missed cleavage sites—2; fixed modification—carbamidomethylation of cysteines; variable modification—oxidation of methionines; precursor mass tolerance filtering—10 ppm; fragment ion mass tolerance—20 ppm; false discovery rate—1% (Supplementary Data 1). The sequence database contained all proteins within the complex. Crosslinking figures were created with XiNet<sup>42</sup> and Xlink Analyzer plugin in Chimera<sup>43</sup>.

**Cryo-EM sample preparation and image processing.** Sample was prepared by mixing pre-annealed DNA scaffold (5'-CAAAGTCACGACCTAGA-CACTGCGAGCTCGAATTCAGTGGAGTGACCTC-3'; 5'-GAGGTCACTC-CAGTGAATTCGAGCTCGCAACAATGAGCACATACCTAGT-3') with core TFIIH, XPA and XPG:E791A in 1.5:1:3:3 molar ratio in final buffer containing 150 mM KCl, 20 mM KOH:HEPES pH 7.0, 10% glycerol, 2 mM MgCl<sub>2</sub>, and 5 mM β-mercaptoethanol. XPG:E791A endonuclease mutant was used to prevent DNA cleavage during the sample preparation<sup>44</sup>. The sample mixture was applied to a sucrose gradient in order to purify the complex from excess factors and fix it with glutaraldehyde<sup>45</sup>. The sucrose gradient was prepared with BioComp Gradient Master 108 (BioComp Instruments) by mixing equal volume of heavy (30% (w/v) sucrose, 150 mM KCl, 2 mM MgCl<sub>2</sub>, 5 mM β-mercaptoethanol, 20 mM KOH:HEPES pH 7.5, and 0.1% glutaraldehyde) and light solutions (10% (w/v) sucrose, 150 mM KCl, 2 mM MgCl<sub>2</sub>, 5 mM β-mercaptoethanol, and 20 mM KOH:HEPES pH 7.5) in 5 ml centrifugation tubes. After 16 h of centrifugation at 4 °C and 175,000 × g, the gradient was fractionated and glutaraldehyde was quenched with lysine (50 mM final) and aspartate (20 mM final). Fractions were dialyzed in Slide-A-Lyzer MINI Dialysis Devices (2 ml and 20 kDa cut-off) (ThermoFisher Scientific) for 10 h against buffer containing 100 mM KCl, 20 mM KOH:HEPES pH 7.5, 1 mM MgCl<sub>2</sub>, 1 mM DTT, 0.5% glycerol (v/v), and 0.004% n-octyl glucoside (w/v). Dialyzed samples were immediately used for cryo-grid preparation. Four microliters of sample was applied to glow-discharged R2/2 gold grids (Quantifoil), which were blotted for 5 s and plunge-frozen in liquid ethane with a Vitrobot Mark IV (FEI) operated at 4 °C and 100% humidity.

Micrographs of the sample were acquired on a FEI Titan Krios G2 transmission electron microscope with a K2 summit direct electron detector (Gatan). Data acquisition was automated with the FEI EPU software package. Micrographs were acquired at a nominal magnification of ×130,000 (1.05 Å/pixel) using a dose rate of 4.55 e<sup>-</sup>/Å<sup>2</sup> per s over the time of 9 s that resulted in a total dose of 41 e<sup>-</sup>/Å<sup>2</sup> fractionated over 40 frames. CTF correction, motion correction, and particle picking was done on-the-fly using Warp<sup>46</sup>. Automated picking in retrained BoxNet implemented in Warp<sup>46</sup> yielded a total of 1,354,997 particles from 8993 micrographs, which were further subjected to two-dimensional (2D) classification in CryoSPARC<sup>47</sup>. After 2D cleaning, 950,000 particles were used for heterogeneous refinement in CryoSPARC. Three ab initio classes obtained from the first 300,000 particles picked during data acquisition were used as an input for the refinement. The class showing clear core TFIIH features was further three-dimensional (3D) classified into six classes using RELION-3<sup>48</sup>. Particles corresponding to the best 3D class were subjected to CTF refinement and Bayesian polishing. Particles were 3D refined and post-processed with automatic B-factor determination in RELION. Final map showed an overall resolution of 3.6 Å according to the gold-standard Fourier shell correlation 0.143 criterion. Owing to flexibility of peripheral regions of core TFIIH, we improved the map quality for five different regions of the complex by focused 3D classification and refinement (processing tree is depicted in Supplementary Fig. 2d). The classifications were performed with particles contributing to the final map without image alignment to speed up the calculations. Masks encompassing the regions of interest were created with UCSF Chimera<sup>49</sup> and RELION. 3D classification of the DNA duplex revealed two alternative DNA conformations within the complex (Supplementary Fig. 2d).

**Model building.** The final cryo-EM map and focused refined maps were used for model building. The final map was denoised in Warp 1.0.6<sup>46</sup>. Structures of ATPase lobes 1 and 2 of XPB, XPD, p44 vWA-like domain, and p52 C-terminus (residues 383–458) from the TFIIH structure (PDB code 5OF4 [<https://doi.org/10.2210/pdb5OF4/pdb>])<sup>6</sup>, as well as the crystal structure of p34 vWA-like domain bound to p44 RING domain (PDB code 5O85 [<https://doi.org/10.2210/pdb5O85/pdb>])<sup>50</sup> were rigid-body fitted into our cryo-EM density in UCSF Chimera<sup>49</sup> and manually adjusted in COOT<sup>51</sup>. Owing to high quality of the EM density, the NTE domain and part of the DRD domain (residues 71–199 and 266–300), as well as the p52 region that interacts with XPB (residues 307–382) were built de novo guided by secondary structure prediction in PSIPRED<sup>52</sup> and bulky amino acid side chains as sequence registers. In case of XPD, we did not observe EM density corresponding to residues 273–325, so we removed this part of the structure. We observed a very strong density for the iron-sulfur cluster indicating that the ligand was not damaged or dissociated during protein expression and purification, as well as sample preparation for cryo-EM. The N-terminal region of p52 (residues 18–289) and zinc-fingers belonging to subunits p34 and p44 were modeled with SWISS-Model<sup>53,54</sup> based on the yeast p52 counterpart (PDB code 5OQJ [<https://doi.org/10.2210/pdb5OQJ/pdb>])<sup>7</sup> and manually adjusted in COOT. Interestingly, the p34 zinc finger region in human contains additional cysteine (C257) and histidine

(H258) residues not present in the yeast counterpart that allows binding of an additional zinc ion. The smallest TFIIH subunit TTDA (p8) was generated in Modeller<sup>55</sup> with the yeast TTDA structure as a reference (PDB code 5OQJ [<https://doi.org/10.2210/pdb5OQJ/pdb>])<sup>7</sup>, rigid-body fitted in our density using UCSF Chimera, and manually adjusted in COOT. The nuclear magnetic resonance structure of truncated human XPA (PDB code 1XPA [<https://doi.org/10.2210/pdb1XPA/pdb>])<sup>56</sup> was also docked in our density as a rigid body and adjusted in COOT. We observed additional helical density that extends from the C-terminus of the docked structure toward the ATPase lobe 2 of XPB when the map is filtered to lower resolution. Secondary structure prediction with PSIPRED shows that residues following the docked XPA C-terminus form a helix, so we extended the C-terminal helix in COOT guided by the cryo-EM density (Supplementary Fig. 3h).

DNA sequence was assigned based on the position of the DNA duplex—single strand junction; however, protein binding to the junction could induce additional DNA melting so register shifts cannot be excluded. DNA duplex was built by docking ideal B-DNA into the density, followed by manual adjustments in COOT. Several rounds of real space refinement and geometry optimization with secondary structure restraints (including base pairing and base stacking restraints) were performed in PHENIX<sup>57</sup>. The DNA duplex—single strand junction and single-strand extensions were manually built in COOT. The EM density for the 5'–3' DNA single strand showed clear separation of sugar, phosphate, and DNA bases for nucleotides A30–G35 and for C40–A41. The decreased quality of EM map for nucleotides A36–A39 and T42, presumably due to increased flexibility of DNA between XPD helicase lobes, allowed the trajectory of DNA to be determined, but the nucleotides were positioned manually in COOT guided by the structure of NS3 helicase in complex with DNA<sup>58</sup> and real space refined in PHENIX. All core TFIIH subunits, XPA, and DNA were first real-space refined in PHENIX separately in their corresponding focused classified maps. Then all components were combined and real-space refined together in the global map. The final model was validated using Molprobity<sup>59</sup> (Supplementary Table 2). Figures were generated using PyMOL (Schrödinger LLC) and UCSF Chimera<sup>49</sup>.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

The cryo-EM density reconstructions were deposited with the Electron Microscopy Data Base (code EMD-4970). The main cryo-EM map is composed of focused refined maps that were deposited under the same accession code together with the globally refined map and all corresponding half maps. The final model was deposited with the Protein Data Bank (code 6RO4). The crosslinking mass spectrometric data have been deposited to the ProteomeXchange Consortium via the PRIDE<sup>60</sup> partner repository with the dataset identifier PXD013947. All other relevant data supporting the key findings of this study are available within the article, its Supplementary Information, and Source Data or from the corresponding author upon reasonable request.

Received: 3 May 2019 Accepted: 28 May 2019

Published online: 28 June 2019

## References

- Schärer, O. D. Nucleotide excision repair in eukaryotes. *Cold Spring Harb. Perspect. Biol.* **5**, a012609 (2013).
- Kraemer, K. H. et al. Xeroderma pigmentosum, trichothiodystrophy and Cockayne syndrome: a complex genotype-phenotype relationship. *Neuroscience* **145**, 1388–1396 (2007).
- Lehmann, A. R. The xeroderma pigmentosum group D (XPD) gene: one gene, two functions, three diseases. *Genes Dev.* **15**, 15–23 (2001).
- Evans, E., Moggs, J. G., Hwang, J. R., Egly, J. M. & Wood, R. D. Mechanism of open complex and dual incision formation by human nucleotide excision repair factors. *EMBO J.* **16**, 6559–6573 (1997).
- Compe, E. & Egly, J. M. Nucleotide excision repair and transcriptional regulation: TFIIH and beyond. *Annu. Rev. Biochem.* **85**, 265–290 (2016).
- Greber, B. J. et al. The cryo-electron microscopy structure of human transcription factor IIH. *Nature* **549**, 414–417 (2017).
- Schilbach, S. et al. Structures of transcription pre-initiation complex with TFIIH and mediator. *Nature* **551**, 204 (2017).
- Abdulrahman, W. et al. ARCH domain of XPD, an anchoring platform for CAK that conditions TFIIH DNA repair and transcription activities. *Proc. Natl Acad. Sci. USA* **110**, E633–E642 (2013).
- Li, C. L. et al. Tripartite DNA lesion recognition and verification by XPC, TFIIH, and XPA in nucleotide excision repair. *Mol. Cell* **59**, 1025–1034 (2015).
- Titov, D. V. et al. XPB, a subunit of TFIIH, is a target of the natural product triptolide. *Nat. Chem. Biol.* **7**, 182 (2011).

11. Fishburn, J., Tomko, E., Galburt, E. & Hahn, S. Double-stranded DNA translocase activity of transcription factor TFIIF and the mechanism of RNA polymerase II open complex formation. *Proc. Natl Acad. Sci. USA* **112**, 3961–3966 (2015).
12. Fan, L. et al. Conserved XPB core structure and motifs for DNA unwinding: implications for pathway selection of transcription or excision repair. *Mol. Cell* **22**, 27–37 (2006).
13. 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 XPC complex and DNA strand scanning. *Mol. Cell* **36**, 642–653 (2009).
14. Mu, D., Wakasugi, M., Hsu, D. S. & Sancar, A. Characterization of reaction intermediates of human excision repair nuclease. *J. Biol. Chem.* **272**, 28971–28979 (1997).
15. He, Y. et al. Near-atomic resolution visualization of human transcription promoter opening. *Nature* **533**, 359–365 (2016).
16. Sugitani, N., Sivley, R. M., Perry, K. E., Capra, J. A. & Chazin, W. J. XPA: a key scaffold for human nucleotide excision repair. *DNA Repair* **44**, 123–135 (2016).
17. Ziani, S. et al. Sequential and ordered assembly of a large DNA repair complex on undamaged chromatin. *J. Cell Biol.* **206**, 589–598 (2014).
18. Coin, F., Oksenyk, V. & Egly, J. M. Distinct roles for the XPB/p52 and XPD/p44 subcomplexes of TFIIF in damaged DNA opening during nucleotide excision repair. *Mol. Cell* **26**, 245–256 (2007).
19. Coin, F. et al. p8/TTD-A as a repair-specific TFIIF subunit. *Mol. Cell* **21**, 215–226 (2006).
20. Yang, Z. et al. Specific and efficient binding of xeroderma pigmentosum complementation group A to double-strand/single-strand DNA0 junctions with 3'- and/or 5'-ssDNA branches. *Biochemistry* **45**, 15921–15930 (2006).
21. Koch, S. C. et al. Structural insights into the recognition of cisplatin and AAF-dG lesion by Rad14 (XPA). *Proc. Natl Acad. Sci. USA* **112**, 8272–8277 (2015).
22. Dubaele, S. et al. Basal transcription defect discriminates between xeroderma pigmentosum and trichothiodystrophy in XPD patients. *Mol. Cell* **11**, 1635–1646 (2003).
23. Liu, H. et al. Structure of the DNA repair helicase XPD. *Cell* **133**, 801–812 (2008).
24. Fan, L. et al. XPD helicase structures and activities: insights into the cancer and aging phenotypes from XPD mutations. *Cell* **133**, 789–800 (2008).
25. Wolski, S. C. et al. Crystal structure of the FeS cluster-containing nucleotide excision repair helicase XPD. *PLoS Biol.* **6**, e149 (2008).
26. Kawaoka, J., Jankowsky, E. & Pyle, A. M. Backbone tracking by the SF2 helicase NPH-II. *Nat. Struct. Mol. Biol.* **11**, 526–530 (2004).
27. Pyle, A. M. Translocation and unwinding mechanisms of RNA and DNA helicases. *Annu. Rev. Biophys.* **37**, 317–336 (2008).
28. Levin, M. K., Gurjar, M. & Patel, S. S. A Brownian motor mechanism of translocation and strand separation by hepatitis C virus helicase. *Nat. Struct. Mol. Biol.* **12**, 429–435 (2005).
29. Mathieu, N., Kaczmarek, N., Ruthemann, P., Luch, A. & Naegeli, H. DNA quality control by a lesion sensor pocket of the Xeroderma pigmentosum group D helicase subunit of TFIIF. *Curr. Biol.* **23**, 204–212 (2013).
30. Sontz, P. A., Mui, T. P., Fuss, J. O., Tainer, J. A. & Barton, J. K. DNA charge transport as a first step in coordinating the detection of lesions by repair proteins. *Proc. Natl Acad. Sci. USA* **109**, 1856–1861 (2012).
31. Brooks, S. C., Adhikary, S., Rubinson, E. H. & Eichman, B. F. Recent advances in the structural mechanisms of DNA glycosylases. *Biochim. Biophys. Acta* **1834**, 247–271 (2013).
32. Coin, F. et al. Nucleotide excision repair driven by the dissociation of CAK from TFIIF. *Mol. Cell* **31**, 9–20 (2008).
33. Ito, S. et al. XPG stabilizes TFIIF, allowing transactivation of nuclear receptors: implications for Cockayne syndrome in XP-G/CS patients. *Mol. Cell* **26**, 231–243 (2007).
34. Greber, B. J., Toso, D. B., Fang, J. & Nogales, E. The complete structure of the human TFIIF core complex. *eLife* **8**, e44771 (2019).
35. Gradia, S. D. et al. MacroBac: new technologies for robust and efficient large-scale production of recombinant multiprotein complexes. *Methods Enzymol.* **592**, 1–26 (2017).
36. Brosey, C. A., Chagot, M. E. & Chazin, W. J. Preparation of the modular multi-domain protein RPA for study by NMR spectroscopy. *Methods Mol. Biol.* **831**, 181–195 (2012).
37. Boehning, M. et al. RNA polymerase II clustering through carboxy-terminal domain phase separation. *Nat. Struct. Mol. Biol.* **25**, 833–840 (2018).
38. Vos, S. M. et al. Architecture and RNA binding of the human negative elongation factor. *eLife* **5**, e14981 (2016).
39. Leclerc, V., Raisin, S. & Léopold, P. Dominant-negative mutants reveal a role for the Cdk7 kinase at the mid-blastula transition in *Drosophila* embryos. *EMBO J.* **19**, 1567–1575 (2000).
40. Sydow, J. F. et al. Structural basis of transcription: mismatch-specific fidelity mechanisms and paused RNA polymerase II with frayed RNA. *Mol. Cell* **34**, 710–721 (2009).
41. Combe, C. W., Fischer, L. & Rappsilber, J. xiNET: cross-link network maps with residue resolution. *Mol. Cell. Proteomics* **14**, 1137–1147 (2015).
42. Yang, B. et al. Identification of cross-linked peptides from complex samples. *Nat. Methods* **9**, 904–906 (2012).
43. Kosinski, J. et al. Xlink Analyzer: software for analysis and visualization of cross-linking data in the context of three-dimensional structures. *J. Struct. Biol.* **189**, 177–183 (2015).
44. Hohl, M., Thorel, F., Clarkson, S. G. & Scharer, O. D. Structural determinants for substrate binding and catalysis by the structure-specific endonuclease XPG. *J. Biol. Chem.* **278**, 19500–19508 (2003).
45. Kastner, B. et al. GraFix: sample preparation for single-particle electron cryomicroscopy. *Nat. Methods* **5**, 53–55 (2008).
46. Tegunov, D. & Cramer, P. Real-time cryo-EM data pre-processing with Warp. *bioRxiv* <https://doi.org/10.1101/338558> (2018).
47. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
48. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).
49. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
50. Radu, L. et al. The intricate network between the p34 and p44 subunits is central to the activity of the transcription/DNA repair factor TFIIF. *Nucleic Acids Res.* **45**, 10872–10883 (2017).
51. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
52. Buchan, D. W. A., Minnici, F., Nugent, T. C. O., Bryson, K. & Jones, D. T. Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res.* **41**, W349–W357 (2013).
53. Biasini, M. et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **42**, W252–W258 (2014).
54. Bordoli, L. et al. Protein structure homology modeling using SWISS-MODEL workspace. *Nat. Protoc.* **4**, 1–13 (2009).
55. Webb, B. & Sali, A. Comparative protein structure modeling using MODELLER. *Curr. Protoc. Protein Sci.* **86**, 2.9.1–2.9.37 (2016).
56. Ikegami, T. et al. Solution structure of the DNA- and RPA-binding domain of the human repair factor XPA. *Nat. Struct. Mol. Biol.* **5**, 701–706 (1998).
57. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
58. Kim, J. L. et al. Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: the crystal structure provides insights into the mode of unwinding. *Structure* **6**, 89–100 (1998).
59. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).
60. Perez-Riverol, Y. et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **47**, D442–d450 (2019).
61. Camenisch, U., Dip, R., Schumacher, S. B., Schuler, B. & Naegeli, H. Recognition of helical kinks by *Xeroderma pigmentosum* group A protein triggers DNA excision repair. *Nat. Struct. Mol. Biol.* **13**, 278–284 (2006).

## Acknowledgements

We thank current and former members of Cramer Laboratory, including C. Bernecky, C. Burzinski, S. Dodonova, H. Hillen, S. Vos, and F. Fischer. G.K. was supported by a Boehringer-Ingelheim PhD Fellowship. G.K. and A.C. have been doctoral students of the Ph.D. program "Molecular Biology" – International Max Planck Research School and the Göttingen Graduate School for Neurosciences, Biophysics, and Molecular Biosciences (GGNB) (DFG grant GSC 226) at the Georg August University Göttingen. P.C. was supported by the Deutsche Forschungsgemeinschaft (SFB860, SPP1935), the European Research Council Advanced Investigator Grant TRANSREGULON (grant agreement no. 693023), and the Volkswagen Foundation.

## Author contributions

G.K. designed and carried out all experiments except for crosslinking mass spectrometry, which was carried out by A.C. D.T. assisted with image processing and C.D. assisted with cryo-EM data acquisition and model building. H.U. supervised mass spectrometry. P.C. designed and supervised research. G.K. and P.C. interpreted the data and wrote the manuscript, with input from all authors.

**Additional information**

**Supplementary Information** accompanies this paper at <https://doi.org/10.1038/s41467-019-10745-5>.

**Competing interests:** The authors declare no competing interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**Peer review information:** *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019