bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The complete structure of the human TFIIH core complex

Basil J. Greber[1,2], Daniel B. Toso[1], Jie Fang[3], Eva Nogales[1,2,3,4,*]

[1] California Institute for Quantitative Biosciences (QB3), University of California, Berkeley, California 94720, USA

[2] Molecular Biophysics and Integrative Bio-Imaging Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

[3] Howard Hughes Medical Institute, University of California, Berkeley, California 94720, USA

[4] Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA

[*] **For correspondence:** Enogales@lbl.gov (E.N.)

Competing interests: The authors declare that no competing interests exist.

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ABSTRACT

Transcription factor IIH (TFIIH) is a heterodecameric protein complex critical for transcription initiation by RNA polymerase II and nucleotide excision DNA repair. The TFIIH core complex is sufficient for its repair functions and harbors the XPB and XPD DNA-dependent ATPase/helicase subunits, which are affected by human disease mutations. Transcription initiation additionally requires the CdK activating kinase subcomplex. Previous structural work has provided only partial insight into the architecture of TFIIH and its interactions within transcription pre-initiation complexes. Here, we present the complete structure of the human TFIIH core complex, determined by phase-plate cryo-electron microscopy at 3.7 Å resolution. The structure uncovers the molecular basis of TFIIH assembly, revealing how the recruitment of XPB by p52 depends on a pseudo-symmetric dimer of homologous domains in these two proteins. The structure also suggests a function for p62 in the regulation of XPD, and allows the mapping of previously unresolved human disease mutations.

INTRODUCTION

Transcription factor IIH (TFIIH) is a 10-subunit protein complex with a total molecular weight of 0.5 MDa that serves a dual role as a general transcription factor for transcription initiation by eukaryotic RNA polymerase II (Pol II), and as a DNA helicase complex in nucleotide excision DNA repair (NER) (Compe & Egly, 2016; Sainsbury et al., 2015). Mutations in TFIIH subunits that cause the inherited autosomal recessive disorders xeroderma pigmentosum (XP), trichothiodystrophy (TTD), and Cockayne syndrome (CS) are characterized by high incidence of cancer or premature ageing (Cleaver et al., 1999; Rapin, 2013). Furthermore, TFIIH is a possible target for anti-cancer compounds (Berico & Coin, 2017) and therefore of great importance for human health and disease.

The TFIIH core complex is composed of the seven subunits XPB, XPD, p62, p52, p44, p34, and p8, and is the form of TFIIH active in DNA repair (Svejstrup et al., 1995), where TFIIH serves as a DNA damage verification factor (Li et al., 2015;

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Mathieu et al., 2013) and is responsible for opening a repair bubble around damaged nucleotides. This activity depends on both the SF2-family DNA-dependent ATPase XPB, and the DNA helicase activity of XPD (Coin et al., 2007; Evans et al., 1997; Kuper et al., 2014). TFIIH function in transcription initiation requires the double-stranded DNA translocase activity of XPB to regulate opening of the transcription bubble (Alekseev et al., 2017; Fishburn et al., 2015; Grünberg et al., 2012), and additionally the CdK activating kinase (CAK) complex, which harbors the kinase activity of CDK7 as well as the Cyclin H and MAT1 subunits (Devault et al., 1995; Fisher et al., 1995; Fisher & Morgan, 1994; Shiekhattar et al., 1995; Svejstrup et al., 1995). Targets of human CDK7 include the C-terminal heptapeptide repeat domain of the largest subunit of Pol II, as well as cell-cycle regulating CDKs (Fisher & Morgan, 1994; Shiekhattar et al., 1995). MAT1 serves as a bridging subunit that promotes CAK subcomplex formation by interacting with Cyclin H and CDK7 (Devault et al., 1995; Fisher et al., 1995), recruits the CAK to the core complex by interactions with XPD and XPB (Abdulrahman et al., 2013; Busso et al., 2000; Greber et al., 2017; Rossignol et al., 1997), and also aids in Pol II-PIC formation by establishing interactions with the core PIC (He et al., 2013; He et al., 2016; Schilbach et al., 2017). The presence of MAT1 inhibits the helicase activity of XPD (Abdulrahman et al., 2013; Sandrock & Egly, 2001), but the mechanism of this inhibition is not fully understood. While the enzymatic activity of XPD is not required for transcription initiation, it is critical for the DNA repair function of TFIIH (Dubaele et al., 2003; Evans et al., 1997; Kuper et al., 2014). Therefore, NER requires the release of the CAK subcomplex from the core complex (Coin et al., 2008). The activities of both XPB and XPD are regulated by interactions with additional TFIIH components, including that of p44 with XPD (Coin et al., 1998; Dubaele et al., 2003; Kim et al., 2015), and those of the p52-p8 module with XPB (Coin et al., 2007; Coin et al., 2006; Jawhari et al., 2002; Kainov et al., 2008). These interactions are likely to be crucial for TFIIH function, as some are affected by disease mutations (Cleaver et al., 1999), but they have been only partially characterized mechanistically.

Our previous structure of the TFIIH core-MAT1 complex at 4.4 Å resolution (Greber et al., 2017) allowed modeling of TFIIH in the best-resolved parts of the

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

density map, but several functionally important regions remained unassigned or only partially interpreted because reliable *de novo* tracing of entire domains in the absence of existing structural models was not possible. Here, we present the complete structure of the human TFIIH core complex in association with the CAK subunit MAT1, determined by phase plate cryo-electron microscopy (cryo-EM) at 3.7 Å resolution. Our structure reveals the complete architecture of the TFIIH core complex and provides detailed insight into the interactions that govern its assembly. Additionally, our cryo-EM maps define the molecular contacts that control the regulation of the XPB and XPD subunits of TFIIH, including the critical p52-XBP interaction, and an extensive regulatory network around XPD, formed by XPB, p62, p44, and MAT1.

RESULTS

Structure determination of TFIIH

To determine the complete structure of the human TFIIH core complex, we collected several large cryo-EM datasets (Supplementary Table 1) of TFIIH immuno-purified from HeLa cells using an electron microscope equipped with a Volta phase plate (VPP) (Danev & Baumeister, 2017) and a direct electron detector camera mounted behind an energy filter. From a homogeneous subset of approximately 140,000 TFIIH particle images identified by 3D classification (Scheres, 2010), we reconstructed a 3D cryo-EM density map at 3.7 Å resolution (Figure 1 – figure supplements 1, 2A-C). This VPP-based cryo-EM map was substantially improved compared to our previous maps obtained without phase plate, both in resolution and interpretability (Figure 1 – figure supplement 2D-G), and enabled building, refinement, and full validation of an atomic model of the TFIIH core complex and the MAT1 subunit of the CAK subcomplex (Figure 1A-C, Figure 1 – figure supplement 2B, C, Supplementary Tables 2, 3), while the remainder of the CAK is invisible in our map due to extreme motion. Tracing and sequence register assignment of protein components modeled *de novo* was facilitated by density maps obtained from focused classification and multibody refinement (Figure 1 – figure supplements 3-5) (Bai et al., 2015; Nakane et al.,

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2018), which resulted in density maps for sub-volumes of the complex at up to 3.6 Å resolution. Both the overall and multibody-refined maps showed clear side chain information (Figure 1 – figure supplement 6A-D). Furthermore, our model was corroborated by existing chemical crosslinking-mass spectrometry (CX-MS) data of human TFIIH (Luo et al., 2015) and site-specific crosslinks from yeast TFIIH (Warfield et al., 2016) (Figure 1 – figure supplement 6E-I).

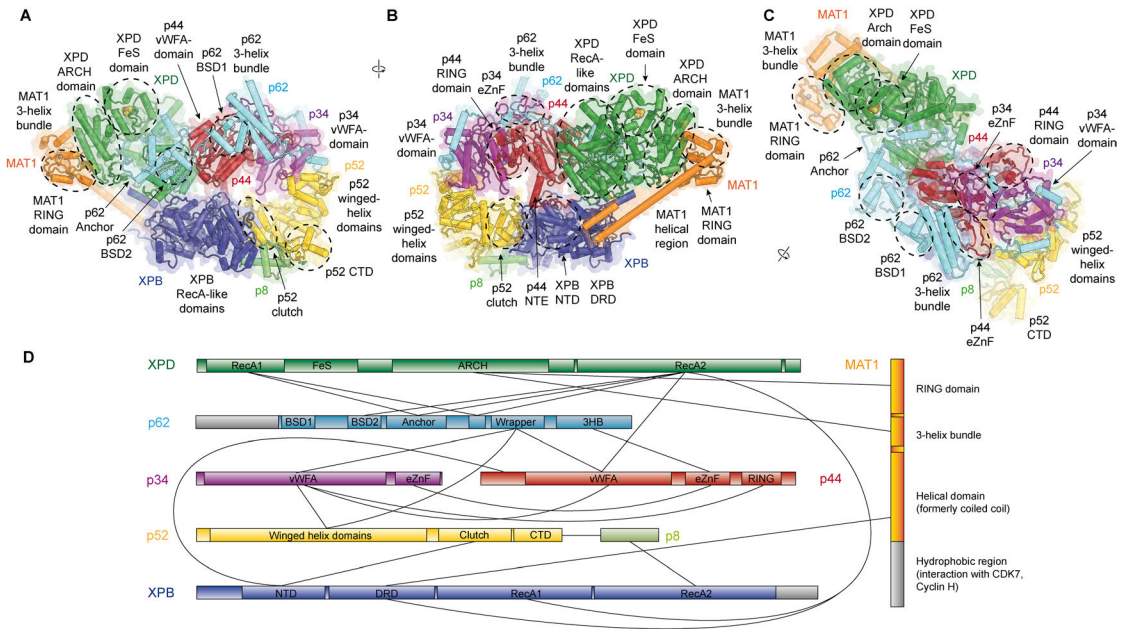


Figure 1. Structure of the TFIIH core complex. (**A**, **B**, **C**) Three views of the structure of the TFIIH core complex and MAT1. Subunits are color-coded and labeled (in color); individual domains are labeled (in black) and circled if needed for clarity. (**D**) Domain-level interaction network between the components of the TFIIH core complex and MAT1 derived from our structure. Proteins are shown with the same colors as in A and major unmodeled regions are shown in grey. Abbreviations: CTD: C-terminal domain; DRD: DNA damage recognition domain; FeS: iron sulfur cluster domain; NTD: N-terminal domain; vWFA: von Willebrand Factor A.

Figure supplement 1. Data processing scheme for global reconstruction.

Figure supplement 2. Resolution estimation, validation statistics, and quality of the density.

Figure supplement 3. Focused classification of the p62 BSD2 domain.

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure supplement 4. Focused classification and interpretation of the MAT1 RING domain density.

Figure supplement 5. Multibody refinement.

Figure supplement 6. Detailed architecture of the TFIIH core complex.

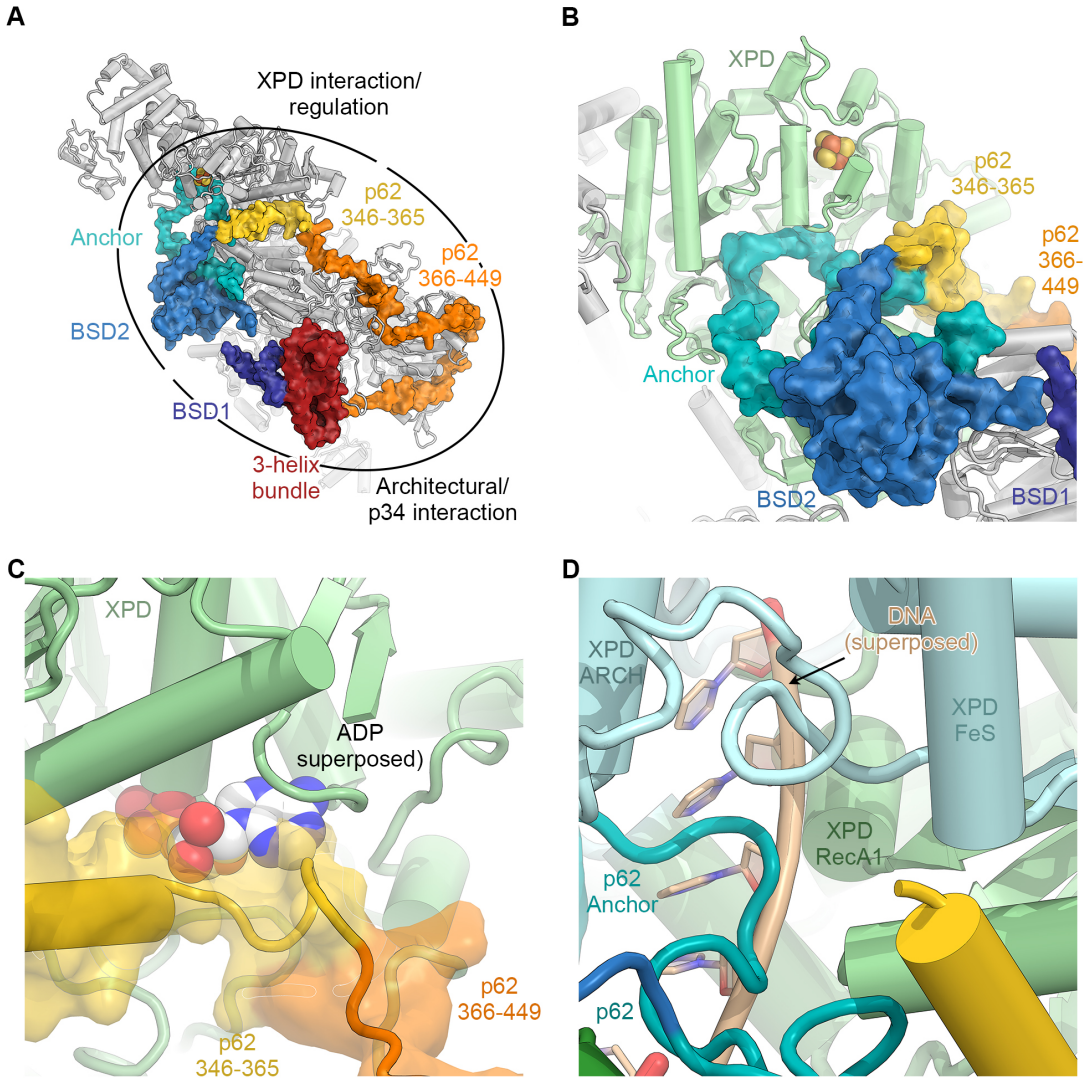
Detailed architecture of TFIIH and structure of p62

Our structure of the TFIIH core complex shows its horseshoe-like overall shape (Figure 1A-C), as observed in previous lower-resolution reconstructions of free and PIC-bound TFIIH (Gibbons et al., 2012; Greber et al., 2017; He et al., 2016; Murakami et al., 2015; Schilbach et al., 2017), and allows us to define the complete set of inter-subunit interactions that lead to the formation of the TFIIH core complex (Figure 1D).

The largest subunits of the complex, the SF2-family DNA-dependent ATPases XPB and XPD, both containing two RecA-like domains (RecA1 and RecA2), interact directly (Greber et al., 2017) and are on one side of the complex, and are additionally bridged by MAT1 (Figure 1B), which has been shown to interact with either ATPase in isolation (Busso et al., 2000). On the side facing away from MAT1, XPD interacts with the von Willebrand Factor A (vWFA) domain of p44 (Coin et al., 1998; Dubaele et al., 2003; He et al., 2016; Kim et al., 2015), which in turn forms a tight interaction with p34 via interlocking eZnF domains (Schilbach et al., 2017) and a p44 RING domain interaction (Radu et al., 2017) (Figure 1B-D, Figure 1 – figure supplement 6J, K), consistent with the formation of a multivalent interaction network between p34 and p44 (Radu et al., 2017). The vWFA domain of p34 recruits p52 by a three-way interaction that involves the most N-terminal winged helix domain in p52 and a helical segment of p62 (Schilbach et al., 2017) (Figure 1 – figure supplement 6L). The p52 C-terminal region comprises two domains; first, the 'clutch' that interacts with XPB (Jawhari et al., 2002) and second, a dimerization module that binds p8 (Kainov et al., 2008), thereby recruiting XPB to TFIIH and cradling XPB RecA2 (see below). In addition to this structural framework that is formed by folded domains, our cryo-EM map reveals several interactions involving extended protein segments,

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

173 including several interactions formed by p62 (Figure 2), and an interaction
174 between the p44 N-terminal extension (NTE) and the N-terminal domain (NTD)
175 of XPB (Figure 1B). To form this interaction, approx. 15 residues of p44 span the
176 distance between the p44 vWFA domain and the XPB NTD, where a small helical
177 motif in p44 contacts XPB residues 72-75, 95-102, and 139-143, in agreement
178 with CX-MS data (Luo et al., 2015) (Figure 1 – figure supplement 6E). Partial
179 deletion of the p44 NTE in yeast causes a slow-growth phenotype, suggesting a
180 functional role for this p44-XPB interaction (Warfield et al., 2016).
181



182
183 **Figure 2. The structure of p62.** (**A**) View of the top surface of the TFIIH core
184 complex; p62 is color-coded by structural region. (**B**) The BSD2 (blue) and XPD
185 anchor segments (teal) of p62 (surface) interact with the region around the XPD
186 substrate-binding cavity. (**C**) Residues 346-365 of p62 (yellow) approach the

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

nucleotide-binding pocket of XPD. ADP superposed from the structure of the DinG helicase (Cheng & Wigley, 2018). (**D**) Superposition of DNA from the structure of the substrate-bound DinG helicase (Cheng & Wigley, 2018) shows that the positions of p62 and RecA1-bound DNA overlap.

Figure supplement 1. Analysis of the structure of p62.

The p62 subunit is almost completely resolved in our structure and exhibits a complex beads-on-a-string-like topology. It fully encircles the top surface of TFIIH (Figures 1C, 2A, Figure 2 – figure supplement 1), interacting with XPD, p52, p44, and p34, in agreement with previous structural findings (Greber et al., 2017; Schilbach et al., 2017). Based on these interactions, p62 can be subdivided into three functional regions: (i) the N-terminal PH-domain, disordered in our structure, is responsible for mediating interactions with components of the core transcriptional machinery (Di Lello et al., 2008; He et al., 2016; Schilbach et al., 2017), transcriptional regulators (Di Lello et al., 2006), and DNA repair pathways (Gervais et al., 2004; Lafrance-Vanasse et al., 2013; Okuda et al., 2017); (ii) residues 108-148 and 454-548 of p62, including the first BSD (BTF2-like, synapse-associated, DOS2-like) domain (BSD1) and the C-terminal 3-helix bundle, play an architectural role by binding to p34 and the extended zinc finger (eZnF) domain of p44 (Figure 2A, C, Figure 1 – figure supplement 6J-L, Figure 2 – figure supplement 1B, C); and (iii) residues 160-365, including the BSD2 domain, are responsible for interactions with and regulation of XPD (Figure 2B-D, Figure 2 – figure supplement 1D-F).

Specifically, p62 residues 160-365 form three structural elements that interact with XPD (Figure 2B), in agreement with previous biochemical, structural, and CX-MS data (Figure 1 – figure supplement 6G) (Jawhari et al., 2004; Luo et al., 2015; Schilbach et al., 2017). First, an α-helix formed by p62 residues 295-318 binds directly to XPD RecA2 and thereby recruits residues 160-258 of p62, comprising the BSD2 domain and adjacent sequence elements, to this surface of XPD RecA2 (Figures 2A, B, Figure 2 – figure supplement 1D). Second, p62 residues 266-287 are inserted into the DNA-binding cavity of XPD (Figures 2B,

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

D), in agreement with previous observations (Schilbach et al., 2017). This inserted p62 segment directly blocks a DNA-binding site on XPD RecA1 (Figure 2D) and localizes near the access path to a pore-like structure between the XPD FeS and ARCH domains. While p62 does not directly contact the DNA-binding surface on XPD RecA2, it may still sterically interfere with DNA binding to the helicase elements of XPD in this region (Figure 2 – figure supplement 1E). Therefore, this segment of p62 may need to move away when XPD binds and unwinds DNA. Third, p62 residues 350-358 form a short α-helix that binds in a cleft between the two RecA-like domains of XPD (Figure 2C), so that it not only closes the entrance to the nucleotide binding pocket in XPD RecA1 (Figure 2 – figure supplement 1F), but also partially overlaps with the predicted location of the nucleotide itself (Figure 2C), strongly suggesting a role for this p62 sequence element in XPD regulation. The density for these structural elements of p62 (residues 260-300 and 346-365) in our cryo-EM map is weaker than for the remainder of the complex, suggesting a dynamic interaction with XPD that enables them to modulate the access to the nucleotide-binding pocket, the DNA-binding cavity, and the DNA-translocating pore of XPD, depending on the functional state of TFIIH. 3D reconstructions of TFIIH classified for these regions of p62 (Figure 1 – figure supplement 3) show globally intact TFIIH, both in the presence and absence of the p62 segments at these XPD sites (Figure 2 – figure supplement 1G-J), supporting our hypothesis of dynamic regulation, rather than the alternative hypothesis of p62 binding to XPD as a requirement for TFIIH stability (Luo et al., 2015).

Molecular basis of XPB recruitment by p52

Our structure of TFIIH resolves the structure and interactions of all four folded domains of human XPB – two RecA-like domains that form the SF2-family type helicase cassette, a DNA damage recognition domain (DRD)-like domain, and an N-terminal domain (NTD) (Figure 3A, Figure 3 – figure supplement 1) – and reveals the molecular basis of XPB recruitment by p52. The XPB NTD encompasses residues 1-165, with the first 55 residues forming an N-terminal extension (NTE), and the remainder assuming a mixed α/β-fold with four α-

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

helices and five β-strands (Figure 3A). The side chain densities in the cryo-EM map (Figure 1 – figure supplement 6A) and CX-MS data (Luo et al., 2015) (Figure 1 – figure supplement 6F) both confirm our assignment of this domain. Existing biochemical data show that the XPB NTD is required for integration of XPB into TFIIH (Jawhari et al., 2002) by forming an interaction with p52 that has been referred to as the 'clutch' (Schilbach et al., 2017). In our structure, the p52 contribution to the clutch encompasses p52 residues 306-399, which, strikingly, assume the same overall fold as the XPD NTD (Figure 3B), as hypothesized previously (He et al., 2016; Luo et al., 2015), thereby forming a pseudo-symmetric dimer of structurally homologous domains. The two domains interact through their β-sheets, via both hydrophobic and charged interactions (Figure 3 – figure supplement 2A-C), and with the most N-terminal β-strand emanating from the XPB NTD extending the p52 β-sheet by additional lateral interactions (Figure 3A, C).

Our structural findings rationalize biochemical data that show that deletion of XPB residues 1-207, but not deletion of residues 1-44, impairs the p52-XPB interaction (Jawhari et al., 2002) (Figure 3C). Our structure is also consistent with data indicating that p52 residues 304-381 are critical for the XPB-p52 interaction (Coin et al., 2007; Jawhari et al., 2002), but does not show any contacts that could explain that reported binding of XPB to p52 residues 1-135 or 1-304 (Jawhari et al., 2002) (Figure 3 – figure supplement 2D, E). The interaction between p52 and XPB not only recruits XPB to TFIIH, but also stimulates its ATPase activity *in vitro* (Coin et al., 2007). Because our structure does not shown any elements of p52 approaching the XPB nucleotide-binding pocket, we propose that this effect is likely induced by the interactions of p52 with the XPB NTD and RecA2, which may, together with p8 (Coin et al., 2006), properly arrange the XPB helicase cassette to bind and hydrolyze ATP (Figure 3D) (Grünberg et al., 2012).

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

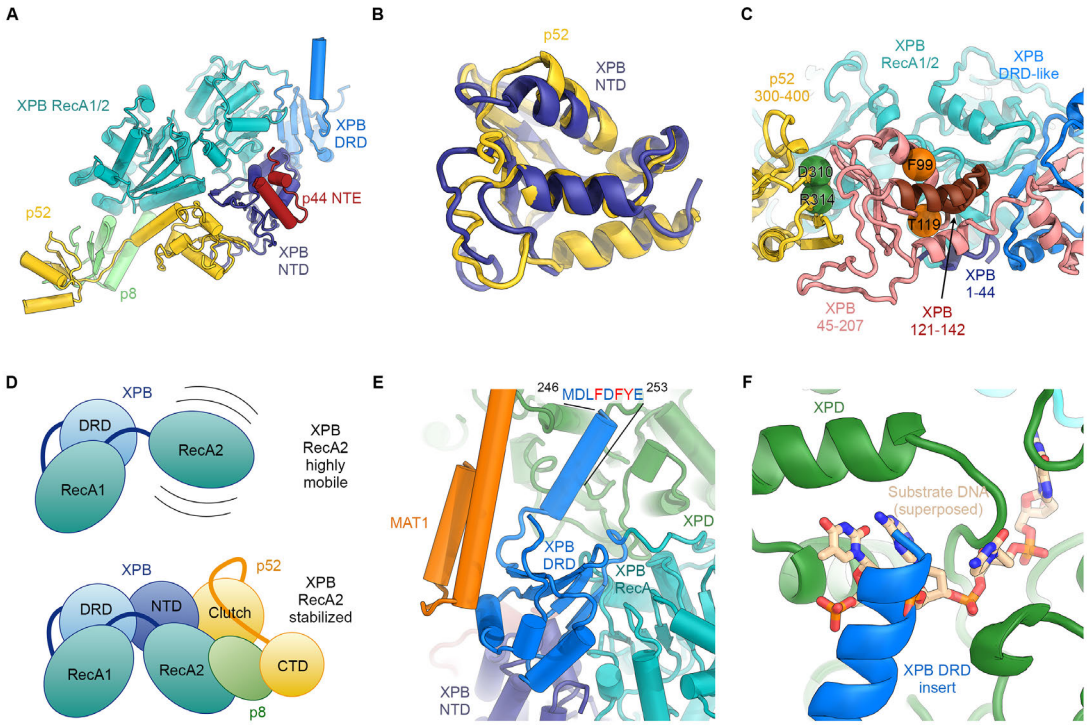


Figure 3. Structure and interactions of XPB. (**A**) Bottom lobe of TFIIH. XPB RecA1/2 teal, DRD blue, NTD dark blue, p52 yellow, p8 green, p44 NTE red. (**B**) Superposition of the XPB NTD and the p52 clutch domain. (**C**) Mapping of mutations on the XPB NTD and the p52 clutch domain; mutated regions are color-coded or shown as spheres (see text for details). (**D**) The combined interactions of the p52 clutch, the p8-p52 CTD dimer, and the XPB NTD with XPB RecA2 may restrict the conformational flexibility of XPB RecA2 to optimize XPB activity. (**E**) An extension of the DRD (blue) contacts XPD (green). The sequence for which formation of an α-helix is predicted (Kelley et al., 2015) is indicated. (**F**) The DRD extension overlaps with the substrate-binding site on XPD RecA2. Substrate DNA modeled from PDB ID 5HW8 (Constantinescu-Aruxandei et al., 2016).

Figure supplement 1. Domain organization of XPB and sequence alignment of the N-terminal regions of XPB and XPB-like enzymes from eukaryotes, archaea, and bacteria.

Figure supplement 2. Structure and interactions of the XPB DRD and NTD.

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The XPB NTD is the site of the two human disease mutations F99S and T119P, which cause XP and TTD, respectively (Cleaver et al., 1999). Or structure shows that neither of these residues is in direct contact with p52 or the RecA-like domains of XPB, suggesting that the F99S and T119P mutations exert their detrimental effects through structural perturbation of the XPB NTD (Figure 3 – figure supplements 1B, 2F-I). Specifically, T119 is located near a turn at the end of a β-strand (Figure 3 – figure supplement 2G), where its side chain points towards the solvent. Nevertheless, this residue is highly conserved in eukaryotic XPB from ciliates to humans, in some archaeal and bacterial XPB homologs (Figure 3 – figure supplement 1B), and in the structurally homologous clutch domain in p52 (Figure 3 – figure supplement 2F-H). This conservation suggests that a threonine at this position is important for the efficient folding of this domain in general, and that the T119P mutation may cause its destabilization, resulting in lower levels of active enzyme in TTD patients. Lower overall levels of properly assembled TFIIH have been shown to be a hallmark of TTD (Botta et al., 2002; Dubaele et al., 2003; Giglia-Mari et al., 2004) and could explain the disease-causing effect of T119P *in vivo* even though recombinant TFIIH carrying this mutation retains some activity in both transcription initiation and NER (Coin et al., 2007). A less likely alternative, given the conservation of the equivalent residue in the p52 clutch, is that T119 is involved in an interaction with a factor that is critical for cellular function, e.g. in NER.

The F99S mutation affects a residue that is conserved throughout eukaryotic XPB (Figure 3 – figure supplement 1B), is inserted into a conserved hydrophobic pocket, and localizes to an α-helix at the XPB contact site with the p44 N-terminal extension (Figure 3 – figure supplement 2I). This mutation is likely to impair the stability and folding of the XPB NTD. Unlike T119P, this mutation leads to impaired DNA opening in NER assays, reduced interaction with p52, reduced ATPase activity (Coin et al., 2007), and strong impairment in DNA damage repair (Riou et al., 1999), suggesting a severe effect on the structure of the XPB NTD.

Natural and synthetic mutations in the *Drosophila melanogaster* homolog of p52 that lead to disease-like phenotypes in flies and have similar defects when

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

introduced into human cells (Fregoso et al., 2007) map directly to the p52-XPB interface, explaining their detrimental phenotypes (Figure 3C, Figure 3 – figure supplement 2A).

Our structure assigns XPB residues 165-300 to a DRD-like domain that connects the NTD to the RecA-like domain (Figure 3A, Figure 3 – figure supplement 1A), the deletion of which is lethal in yeast (Warfield et al., 2016). The DRD is a DNA-binding domain found in DNA repair enzymes and chromatin remodelers (Mason et al., 2014; Obmolova et al., 2000) and has been implicated in DNA damage recognition in archaeal XPB (Fan et al., 2006; Rouillon & White, 2010). Our 3.7 Å-resolution map of TFIIH reveals that in eukaryotic XPB, one β-strand of the DRD of archaeal XPB is replaced by an insertion of approximately 70 residues that exhibits relatively low sequence conservation (Figure 3E, Figure 3 – figure supplement 1B) and shifts the domain boundaries of the human XPB DRD-like domain with respect to previous sequence alignments (Fan et al., 2006; Oksenych et al., 2009). The part of this insertion resolved in our map consists of a negatively charged linker and an α-helical element that contacts XPD directly (Figure 3E, F). The surface on XPD involved in this interaction has been implicated in the initial step of DNA substrate binding by XPD (Constantinescu-Aruxandei et al., 2016; Kuper et al., 2012). Density features and secondary structure prediction indicate the presence of several aromatic side chains of XPB near the interface (Figure 3E), where they might form contacts resembling those of nucleoside bases of XPD-bound DNA substrates (Figure 3F). Thus, XPB may modulate substrate binding by XPD, further reinforcing the idea that XPD activity is regulated by several other components of TFIIH.

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

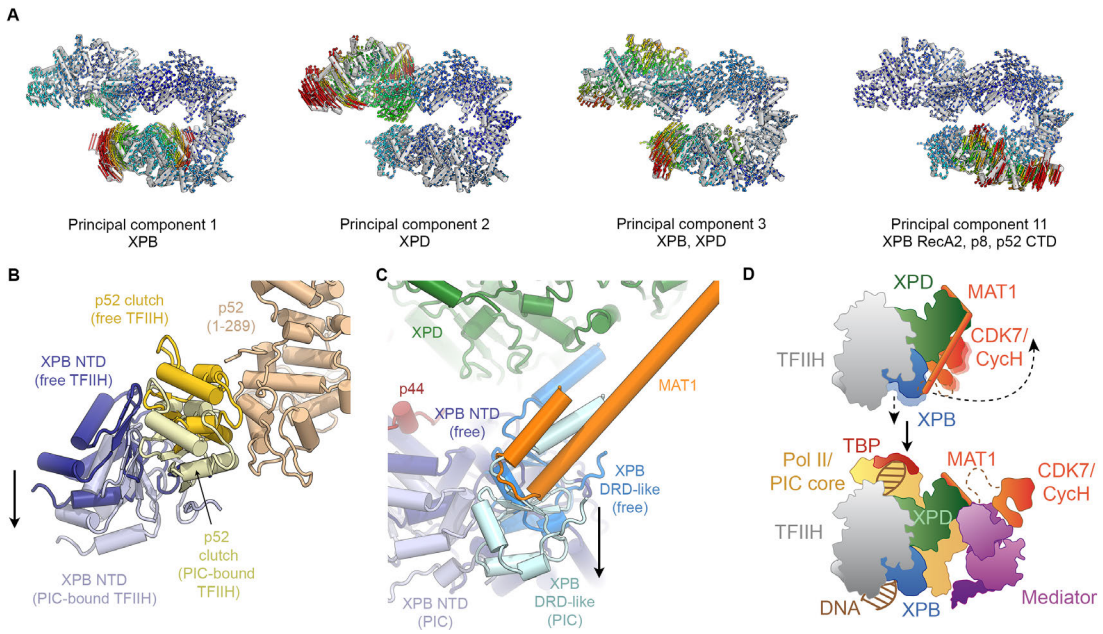


Figure 4. Conformational dynamics of TFIIH. (**A**) Results of multibody analysis. Several major modes of motion involve the enzymatic subunits of the TFIIH core complex or their domains. (**B**) Motion of the p52 clutch domain (closed conformation gold, open conformation light yellow) and associated XPB NTD (closed conformation blue, open conformation light blue) relative to the remainder of p52 (brown), based on comparison of free and PIC-bound TFIIH structures. (**C**) XPB motions from the closed conformation (free TFIIH; darker blue hues) and the open conformation (TFIIH-PIC; lighter hues). The MAT1-XPB contact dissociates during this rearrangement. (**D**) Schematic model for the conformational transitions in MAT1 and repositioning of the CAK kinase module during Pol II-PIC entry of TFIIH.

Figure supplement 1. Analysis of the conformational variance of TFIIH.

Figure supplement 2. Conformational dynamics of TFIIH and comparison with the PIC-bound TFIIH.

Conformational dynamics of the TFIIH core complex

In order to investigate the dynamics of TFIIH, we analyzed the conformational landscape of the particles in our cryo-EM dataset (Figure 4A, Figure 4 – figure supplement 1) (Nakane et al., 2018). This analysis revealed the relative motions

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

of the two ATPases and their domains (Figure 4A). The major mode of motion, which involves the breaking of the interaction between XPB and XPD, resembles the conformational transition of TFIIH when it enters the Pol II-PIC and binds to DNA (Greber et al., 2017; He et al., 2016; Schilbach et al., 2017) (Figure 4 – figure supplement 2A-C). Analysis of our TFIIH structure and comparison with that of the complex within the Pol II-PIC maps (He et al., 2016; Schilbach et al., 2017) allowed us to identify a specific rearrangement at the interface between the clutch and adjacent winged helix domain in p52 (Figure 4B) as the basis of this conformational change. A structural unit composed of XPB, p8 and the clutch domain of p52 undergoes a downward motion upon DNA-binding within the Pol II-PIC (Figure 4B, Figure 4 – figure supplement 2B, C). This conformational change in TFIIH upon PIC entry also appears to break the interaction between MAT1 and the XPB DRD-like domain (Figure 4C), which in turn might serve to enable positioning of the CDK7-cyclin H dimer within the CAK subcomplex at the appropriate location for Pol II-CTD phosphorylation in the mediator-bound Pol II-PIC (Figure 4D) (Schilbach et al., 2017). Our structural comparison also reveals that a TFIIE-XPB interaction that has been implicated in XPB regulation (Schilbach et al., 2017) may depend on the existence of the open conformation of TFIIH, as there would be steric hindrance in a complex involving the closed conformation of TFIIH (Figure 4 – figure supplement 2D-F).

Structure of XPD

The structure of XPD shows the conserved domain arrangement of two RecA-like domains (RecA1 and RecA2), with the FeS and ARCH domain insertions in RecA1 (Constantinescu-Aruxandei et al., 2016; Fan et al., 2008; Kuper et al., 2012). The quality of the map allowed us to interpret the density for the N- and C-termini of XPD, which closely approach each other near the nucleotide-binding site within RecA1 (Figure 5A). The N-terminus of XPD forms a short two-stranded β-sheet near the nucleotide-binding site in XPD RecA1 (which is empty in our structure) and may contribute to the stabilization of the bound nucleotide via the aromatic side chains of Y14 and Y18, the latter being affected by the Y18H mutation in an XP/TTD patient (Kralund et al., 2013) (Figure 5 – figure supplement 1A-C). The

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

XPD C-terminal segment runs along the side of XPD RecA2 and interacts with the linker between RecA1 and RecA2 (Figure 5 – figure supplement 1D). The C-terminal segment includes the site of the K751Q polymorphism (Figure 5 – figure supplement 1D), and deletion of this terminal segment causes XP in human patients (Cleaver et al., 1999).

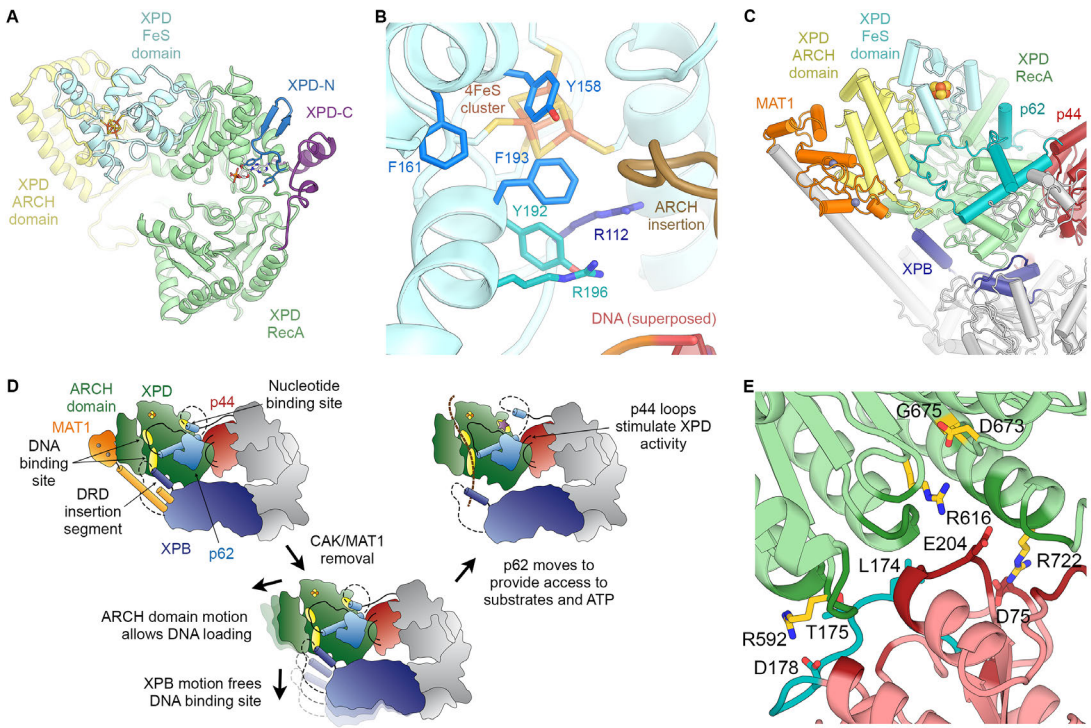


Figure 5. Structure and regulation of XPD. (**A**) Structure of XPD colored by domain. N- and C-terminal segments (blue and purple, respectively) of XPD are indicated. An ADP molecule (superposed from the structure of DinG (Cheng & Wigley, 2018)) denotes the nucleotide-binding pocket in XPD RecA1, which is empty in our structure. (**B**) Structure of the FeS domain. Residues critical for XPD enzymatic activity (blue) and damage verification (teal) are indicated. The R112H mutation causes TTD in human patients. ARCH domain insertion brown. DNA superposed from (Cheng & Wigley, 2018). (**C**) Interaction network of XPD with surrounding TFIIH subunits (interacting regions colored, remainder grey). (**D**) Cartoon model for repression and de-repression of XPD by MAT1, XPB, and p62. (**E**) XPD-p44 interacting regions (defined as residues within < 4 Å of the neighboring protein) are colored in dark green (XPD) and dark red (p44). Residues discussed in the text are shown as sticks; those with mutation data

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(natural variants or experimental constructs) are colored yellow on XPD, teal on p44. The remainder of the β4-α5 loop harboring the synthetic p44 mutations is colored teal as well.

Figure supplement 1. Structure of XPD.

Figure supplement 2. XPD-p44 interaction.

Before XPD-bound DNA reaches the helicase motifs in the RecA like domains, it passes through a pore-like structure next to the 4FeS cluster at the interface between the FeS and ARCH domains (Figure 5 – figure supplement 1E, F) (Cheng & Wigley, 2018; Constantinescu-Aruxandei et al., 2016; Kuper et al., 2012; Liu et al., 2008; Wolski et al., 2008). This region was poorly defined in previous TFIIH reconstructions, but our cryo-EM map now shows side-chain densities for the aromatic residues Y158, F161, and F193, which are critical for the DNA-binding, ATPase, and helicase activities of XPD (Kuper et al., 2014), as well as for residues Y192 and R196, which form part of a DNA lesion recognition pocket (Mathieu et al., 2013) (Figure 5B). This functionally important region is only partially conserved in archaeal XPD homologs (Figure 5 – figure supplement 1G-I) (Fan et al., 2008; Kuper et al., 2012; Wolski et al., 2008). A eukaryotic-specific loop insertion in the XPD ARCH domain (Greber et al., 2017; Schilbach et al., 2017) closely approaches this binding pocket (Figure 5B) and may serve to regulate the binding of DNA in the lesion recognition pocket such as to prevent untimely access of substrates to the XPD pore.

Interactions and regulation of XPD

Our structure of TFIIH shows that XPD forms architectural and regulatory interactions with four other TFIIH subunits: XPB, p62, p44, and MAT1, which together form a cradle-like structure around XPD (Figure 5C). We described above two interactions that could potentially regulate XPD activity: the newly defined interaction of an insertion element in the XPB DRD with a DNA-binding site in XPD (Figure 3E, F); and XPD-p62 interactions involving the XPD nucleotide binding pocked and DNA binding cavity (Figure 2, Figure 2 – figure

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

supplement 1D-F) that implicate p62, as well as XPB, in XPD regulation. Additionally, it is known that the helicase activity of XPD is inhibited by the CAK subcomplex (Araújo et al., 2000; Sandrock & Egly, 2001). The contacts we see between MAT1 and XPD localize to the ARCH domain of XPD and the N-terminal RING domain and helical bundle of MAT1 (residues 1-210) (Figure 5C), in agreement with previous structural (Greber et al., 2017; Schilbach et al., 2017) and biochemical analysis (Abdulrahman et al., 2013; Luo et al., 2015; Warfield et al., 2016). The interaction between the XPD ARCH domain and the MAT1 helical bundle is characterized by charge complementarity (Figure 5 – figure supplement 1J-M). This interface is highly dynamic, enabling the release of MAT1 and the entire CAK subcomplex from TFIIH during NER, as well as its subsequent re-association to regenerate a transcription-competent TFIIH (Coin et al., 2008).

Insertion of substrate DNA into the pore between the XPD ARCH and FeS domains requires the flexibility of the XPD ARCH domain (Constantinescu-Aruxandei et al., 2016), and large domain motions have been observed in the structure of the DNA-bound homologous helicase DinG upon nucleotide binding (Cheng & Wigley, 2018). This suggests a role for the mobility of the ARCH domain in both DNA loading and DNA translocation by the XPD helicase. Our structure suggests that binding of the MAT1 helical bundle and RING domain to the ARCH domain may prevent such motion and therefore the subsequent substrate loading and XPD helicase activity (Figure 5D), in agreement with biochemical data that show XPD inhibition upon MAT1 binding (Sandrock & Egly, 2001), as well as reduced single-stranded DNA affinity of TFIIH in the presence of the CAK (Li et al., 2015). Conversely, release of MAT1 from XPD might allow the ARCH domain to move more freely, thereby de-repressing XPD. Furthermore, displacement of the MAT1 α-helix that connects XPD to XPB may allow XPB to move away from XPD, thereby unmasking the substrate-binding site on XPD RecA2 that is otherwise occluded by the DRD insertion element (Figure 5D). This latter conformational change would be similar, overall, to that seen for TFIIH upon incorporation into the Pol II-PIC, where XPD and XPB move apart and density for the MAT1 helix is missing (Figure 4C) (He et al., 2016; Schilbach et al., 2017). We propose that the combined unmasking of the XPD

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

substrate binding site and the enhanced flexibility of the XPD ARCH domain may both contribute to de-repression of the XPD helicase upon release of MAT1. This mechanism of XPD inhibition by MAT1 does not exclude the possibility of additional repression of NER activity by the CAK subcomplex through phosphorylation of NER pathway components (Araújo et al., 2000).

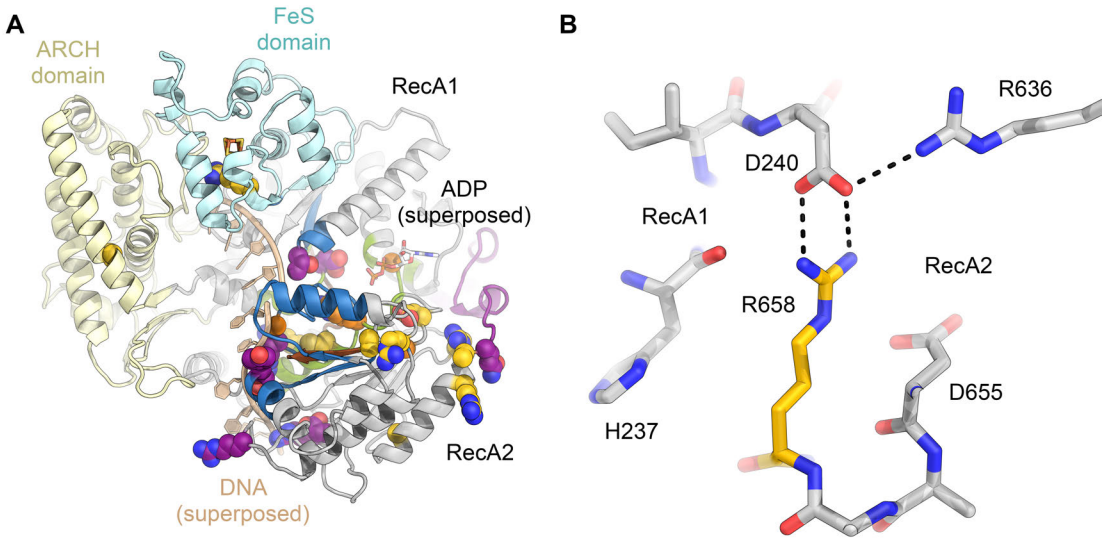


Figure 6. Disease mutations in XPD. (**A**) Residues affected by disease mutations are shown as spheres (XP purple; TTD yellow; CS-XP orange). Conserved helicase elements for DNA binding are shown in blue, for nucleotide binding and hydrolysis in green, and for coupling of nucleotide hydrolysis and DNA translocation in brown. DNA superposed from (Cheng & Wigley, 2018). (**B**) Salt bridge between R658 (RecA2) and D240 (RecA1) visualized in our structure that is affected by the temperature sensitive TTD mutation R658C (Vermeulen et al., 2001).

Figure supplement 1. Mapping of XPD disease mutations on the TFIIH structure.

Our structure also resolves in detail the XPD-p44 interaction, a known regulatory interface (Dubaele et al., 2003; Kim et al., 2015; Kuper et al., 2014) affected by numerous disease mutations (Cleaver et al., 1999; Greber et al., 2017; Kuper et al., 2014) (Figure 5E). The relatively small interaction surface between p44 and XPD, of just 1010 Å^2 (Figure 5 – figure supplement 2A), contrasts with the much

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

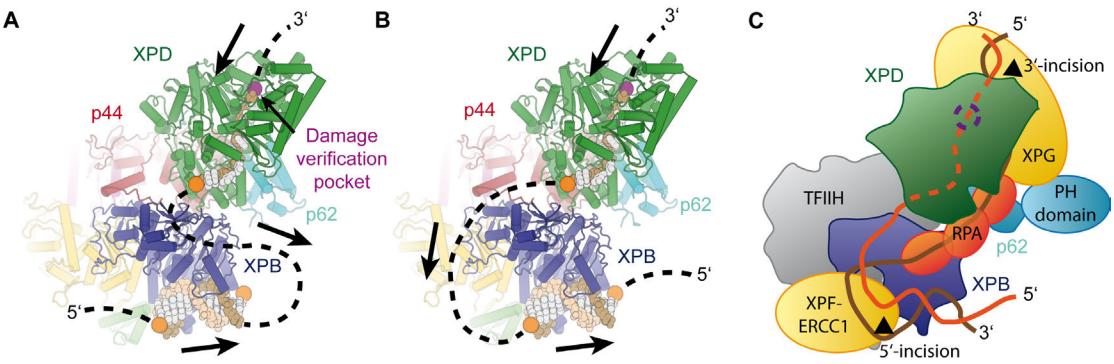
larger buried surface of 3370 Å^2 between XPD and p62, or 1570 Å^2 for the p52-XPB interaction. This smaller interaction surface may result in higher sensitivity to mutations that localize at the XPD-p44 interface. Our structure, thus, rationalizes the deleterious effect of a number of natural and synthetic mutations in this interface (Supplementary Note 1, Figure 5 – figure supplement 2B-E), including mutations L174W and T175R in the β4-α5 loop of p44 (Figure 5E) (Kim et al., 2015; Seroz et al., 2000), which may lead to steric clashes in the densely packed interface (Figure 5 – figure supplement 2B), and the XPD R722W mutation (Kuper et al., 2014), which disrupts the salt bridge with D75 in p44 and may additionally cause steric clashes with neighboring p44 residues due to the bulky tryptophan side chain (Figure 5 – figure supplement 2C). Our structure also shows that, in contrast to a previously proposed model (Luo et al., 2015), the XPD R616P, D673G, and G675R disease mutations act either via disruption of the XPD structure or the XPD-p44 interface, but not via disruption of the interaction with p62 (see Supplementary Note 1, Figure 5 – figure supplement 2D). Notably, the p44-dependet stimulation of XPD activity does not depend on the presence of p62 (Dubaele et al., 2003; Kim et al., 2015; Kuper et al., 2014). We were also able to map a number of disease mutations onto our XPD structure (Figure 6A) and analyze in detail the interactions involving the affected residues (example shown in Figure 6B). Our analysis confirms that XP mutations mostly localize near the helicase substrate-binding or active sites, while TTD mutations predominantly localize to the periphery of XPD (Figure 6, Figure 6 – figure supplement 1) (Fan et al., 2008; Liu et al., 2008), where they disrupt TFIIH assembly and cause the transcription defects that are a hallmark of this disease (Dubaele et al., 2003) (Supplementary Note 2).

DISCUSSION

Our study reveals the complete structure of the TFIIH core complex and provides mechanistic insights into the regulation of its two component helicases. Specifically, it shows XPD wrapped by numerous interactions with XPB, p62, p44, and MAT1 (Figure 5C, D), indicating how its activity can be tightly controlled and de-repressed only when its enzymatic function is needed. XPD activity is not

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

558 needed and most likely inhibited during transcription initiation, but it may also
559 be tightly controlled during NER, when repair bubble opening and lesion
560 verification need to be coordinated with the recruitment and activation of the
561 damage recognition and processing machinery (Figure 7 and Discussion). While
562 the regulation of XPD by MAT1 and p44 has been studied in some detail, and the
563 domain motions in TFIIH suggest a straightforward mechanism for liberating the
564 substrate-binding site on XPD RecA2, less was known about the interplay
565 between XPD and p62. Our structure now shows how p62 is able to impede both
566 substrate and nucleotide binding in XPD RecA1, and hints at dynamic structural
567 changes of p62 during de-repression and enzymatic activity of XPD, possibly
568 regulated by other components of the transcription or NER pathways.
569



570
571 **Figure 7. Implications for assembly of the repair bubble during NER.** (**A**)
572 Schematic of DNA-bound TFIIH (DNA damage verification pocket in XPD and
573 DNA 5'-phosphates indicated by purple and orange spheres, respectively)
574 Binding of both XPB and XPD to the damaged strand would lead to extrusion of a
575 bubble when XPD scans in the 5'-3' direction, while XPB may be stationary or
576 contribute to bubble extrusion if translocating in the 3'-5' direction (DNA
577 superposed from PDB IDs 6FWR, 5OQJ (Cheng & Wigley, 2018; Schilbach et al.,
578 2017)). (**B**) Binding of XPB to the undamaged strand would enable the entire
579 complex to scan in 5'-3' direction, given the opposing polarities of the two
580 ATPases/helicases involved. (**C**) Model for the assembled repair bubble.
581 Positions of NER factors are approximate. XPG-p62 PH domain interaction
582 according to (Gervais et al., 2004). See Discussion for details.
583

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Our results allow us to put extensive biochemical data on the NER pathway into a structural context (Figure 7). Depending on whether XPB binds to the damaged (Figure 7A) or undamaged (Figure 7B) strand, the combined action of XPD and XPB could lead to the extrusion of a DNA bubble (Figure 7A) or to the tracking of the entire complex towards the lesion (Figure 7B), which is initially located 3' of the TFIIH binding site (Sugasawa et al., 2009). The latter hypothesis is attractive in the context of biochemical data that show that XPD tracks along the damaged strand in the 5' to 3' direction until it encounters the DNA lesion in order to verify the presence of a bona fide NER substrate (Buechner et al., 2014; Li et al., 2015; Mathieu et al., 2013; Naegeli et al., 1993; Sugasawa et al., 2009; Wirth et al., 2016). It is worth noting that the length of DNA fragments excised during NER is approx. 29 nt, with 22 nt located 5' and 5 nt located 3' of a thymine dimer lesion (Huang et al., 1992). According to our structure, the 22 nt 5'-fragment corresponds well to the estimated 20 nt of DNA that are required to span the distance from the DNA damage verification pocket in XPD (Mathieu et al., 2013) to the helicase elements of XPB. This proposal is compatible with a model in which TFIIH sitting on the open repair bubble might track towards the lesion, where it would stop due to inhibition of XPD (Li et al., 2015; Mathieu et al., 2013; Naegeli et al., 1993), at which point double incision could be initiated. However, this model (Figure 7B) would require strong DNA bending before both XPB and XPD could be loaded. Additionally, it has not been fully resolved whether XPB participates in DNA translocation or unwinding during TFIIH activity in NER (Li et al., 2015), which would be required in the tracking model (Li et al., 2015), or whether it exclusively acts to anchor the complex in the vicinity of the DNA lesion (Coin et al., 2007; Oksenych et al., 2009).

Independently of the orientation of the repair bubble, our structural data are compatible with literature data introduced above and a model (Figure 7C) that localizes XPG near XPD and p62 (site of 3'-incision), XPF-ERCC1 near XPB (site of 5'-incision), and with RPA binding the non-damaged strand (Fagbemi et al., 2011). We have currently not included XPA in this model because its interactions with distinct partners or participation in various processes, such as involvement in CAK release (Coin et al., 2008), binding to p8 (Ziani et al., 2014), and

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

participation in helicase stalling after lesion recognition (Li et al., 2015), suggest its localization to various, often distant sites on TFIIH, or the repair bubble in general (Sugitani et al., 2016).

In summary, our structure of the human TFIIH core complex reveals the interactions that govern the architecture and function of this molecular machine, provides new insights into the regulation of its enzymatic subunits, and thus constitutes an excellent framework for further mechanistic studies of TFIIH in the context of larger DNA repair and transcription assemblies.

METHODS

TFIIH purification, cryo-EM specimen preparation, and data collection

TFIIH was purified and cryo-EM grids were prepared on carbon-coated C-flat CF 4/2 holey carbon grids (Protochips) using a Thermo Fisher Scientific Vitrobot Mk. IV, as previously described (Greber et al., 2017). To improve on our previous 4.4 Å cryo-EM map of human TFIIH (Greber et al., 2017), which was based on 4 cryo-EM datasets (3 of which were retained in the 4.4 Å reconstruction, datasets 8-10 in Supplementary Table 1) from a Titan microscope (Thermo Fisher Scientific) equipped with a side-entry holder (Gatan) and a K2 Summit direct electron detector (Gatan), we collected new data (dataset 7 in Supplementary Table 1) on a Titan KRIOS microscope (Thermo Fisher Scientific) operated at 300 kV extraction voltage and equipped with a C_S-corrector, a K2 Summit direct electron detector (Gatan) operated in super-resolution counting mode, and a Quantum energy filter (Gatan). This dataset was collected under the same imaging conditions as our previous data (i.e. 37,879 x magnification resulting in 1.32 Å pixel size, and at a total exposure of 40 e^- $Å^{-2}$), except for the change of microscope. Datasets 7-10 could be combined to yield a cryo-EM map at 4.3 Å resolution (not shown), however, this did not lead to a substantial improvement in map quality, suggesting that particle alignment quality was limiting. We therefore opted to collect further data on a Titan KRIOS electron microscope (Thermo Fisher Scientific) operated at 300 kV acceleration voltage and equipped

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

with a Volta Phase Plate (VPP), a Gatan Quantum energy filter (operated at 20 eV slit width), and a Gatan K2 Summit direct electron detector (operated in super-resolution counting mode). VPP data (datasets 1-6 in Supplementary Table 1) were collected according to the defocus acquisition technique (Danev & Baumeister, 2017; Khoshouei et al., 2017) at 43,478 x magnification, resulting in a physical pixel size of 1.15 Å on the object scale, with a total electron exposure of 50 e^- $Å^{-2}$ at an exposure rate of 6.1 e^- $Å^{-2}$ s^{-1} during an exposure time of 8.25 s, dose fractionated into 33 movie frames (50 frames for dataset 6). Data collection was monitored on-the-fly using FOCUS (Biyani et al., 2017) to ensure proper evolution of the VPP-induced phase shift.

Cryo-EM data processing

Initially, we used data collected in 10 microscopy sessions, 6 sessions using the VPP and 4 sessions without the VPP, resulting in > 30'000 total micrographs, of which approx. 15,000 were retained after inspection of the quality of Thon rings and CTF fitting (for details, see Supplementary Table 1). Movie stacks were aligned and dose weighed using MOTIONCOR2 (Zheng et al., 2017). The aligned, dose weighed sums from the datasets collected at 1.32 Å pixel size (datasets 7-10) were up-sampled to 1.15 Å per pixel to match the scale of the micrographs collected using the VPP (datasets 1-6) after calibrating the two magnifications to each other based on 3D reconstructions computed from the two types of data. CTF parameters were estimated using GCTF (Zhang, 2016) and particles were picked using GAUTOMATCH (Kai Zhang, MRC Laboratory of Molecular Biology, Cambridge UK) or RELION (Scheres, 2015) using templates generated from a preliminary run without reference templates. All subsequent data processing was performed in RELION 2 (Kimanius et al., 2016; Scheres, 2012) or RELION 3 (Nakane et al., 2018; Zivanov et al., 2018).

To remove false positive particle picks and broken particles, initial 3D classification was performed on each dataset individually (datasets 3, 4, 5), or on a few pooled datasets if appropriate (datasets 1 and 2 were pooled as they used the same batch of specimen; the non-VPP datasets 7-10 were joined because only

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

few micrographs were retained due to more stringent quality criteria compared to our previous study; and dataset 6 was initially classified together with particles from dataset 4 to compensate for particle orientation bias, see Figure 1 – figure supplement 1). In summary, a total of > 2,000,000 initial particle picks were subjected to this initial 3D classification, identifying approx. 820,000 intact particles that were subjected to further processing. After 3D auto-refinement and another round of 3D classification, performed separately for the VPP and non-VPP data because the two data types were spuriously separated into distinct classes in combined RELION 3D classification runs, the best classes (one from VPP and non-VPP data each) resulting from the high-resolution 3D classifications were refined according to the gold-standard refinement procedure (fully independent half-sets), resulting in a 3.9 Å-resolution reconstruction according to the FSC = 0.143 criterion (Rosenthal & Henderson, 2003; Scheres & Chen, 2012). Beam tilt refinement in RELION 3 (Zivanov et al., 2018) improved the map computed from the VPP data to 3.7 Å resolution. The non-VPP data no longer improved the reconstruction after beam tilt correction and was therefore discarded at this point. The final map was post-processed by application of b-factor of -175 A^2 and low-pass filtration to the nominal 3.7 Å resolution for visualization and later coordinate refinement.

To facilitate the interpretation of less ordered or only partially occupied parts of the structure, including the p62 BSD2 domain, the MAT1 RING domain, the MAT1 three-helix bundle at the XPD arch domain, and the N-terminus of XPD, we used signal subtracted classification (Bai et al., 2015; Nguyen et al., 2015) (p62 BSD2 domain, Figure 1 – figure supplement 3), focused classification (MAT1 RING domain, Figure 1 – figure supplement 4), and multibody refinement (Nakane et al., 2018) (MAT1 three-helix bundle and XPD N-terminus, Figure 1 – figure supplement 5). For these classification procedures, we used only the VPP data because 3D classification separated VPP and conventional cryo-EM data into distinct classes, rendering combined classification ineffective. Multibody refinement led to only a slight improvement in resolution (to 3.6 Å) relative to the overall refined best map, and only during the first two iterations, likely due to the relatively small size of the individual bodies and the resulting limited

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

signal for alignment. However, the above-mentioned structural elements showed improved density features (Figure 1 – figure supplement 1B) and could be more reliably interpreted in the multibody-refined XPD-MAT1 map (green in Figure 1 – figure supplement 5; also used in Figure 1 – figure supplement 2E, F). Overall, the use of VPP data in this work resulted in substantial improvements both in nominal resolution (Figure 1 – figure supplement 2D) and map quality (Figure 1 – figure supplement 2E-G) compared to our previous 4.4 Å-resolution structure.

Model building and refinement

The previous structure of the human TFIIH core complex (Greber et al., 2017) and of yeast TFIIH in the Pol II-PIC (Schilbach et al., 2017) were docked into the cryo-EM map and used as the basis for atomic modeling in O (Jones et al., 1991) and COOT (Emsley et al., 2010). In addition to these models, the structure of the human p34 VWFA-p44 RING domain complex (Radu et al., 2017), the N-terminal RING domain of MAT1 (Gervais et al., 2001), the C-terminal RecA-like domain of human XPB (Hilario et al., 2013) and several homology models for the p52 winged-helix domains generated using the PHYRE2 web server (Kelley et al., 2015) based on templates PDB IDs 3F6O and 1STZ (Liu et al., 2005) were used for model building.

Building of 52: The structure of p52 was traced and built completely *de novo*, with the exception of the winged-helix domains and the very C-terminal domain, where a homology model was placed into the cryo-EM map together with the p8 structure (Kainov et al., 2008; Vitorino et al., 2007) and adjusted to the density.

XPB: The structure of the XPB NTD was also built *de novo*, the structure of the DRD was extensively re-built, and the RecA-like domains were rebuilt to improve the fit to the density.

XPD: The improved cryo-EM map enabled detailed re-building of XPD, including correction of register shifts in the more poorly ordered regions of the protein and extension of the termini.

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

MAT1: The N-terminal MAT1 RING domain (Gervais et al., 2001) was first docked into the focused classified density and combined with the rest of the TFIIH model, were it helped guide the assignment of the sequence register to the MAT1 model, in combination with density features of large side chains in the helical regions.

Building of p34: The human p34 structure (Radu et al., 2017) was docked into the map as is, extended near the interaction site with p52, and combined with a completely re-built model of the C-terminal eZnF domain (Schilbach et al., 2017).

Building of p44: The p44 VWFA fold needed only minor rebuilding and was combined with the eZnF domain and the C-terminal human RING domain model (Radu et al., 2017; Schilbach et al., 2017). The p44 NTE was built according to the density at the contact site with XPB and guided by CX-MS data (Luo et al., 2015). Both the features of the cryo-EM map and crosslinks of the p44 NTE to p34, p52, and XPB (Figure 1 – figure supplement 6E) unambiguously confirm the tracing of this segment towards the XPB NTD, rather than alternative tracing towards XPD (this density is now assigned to p62, in agreement with p62-XPD crosslinks; see below and Figure 1 – figure supplement 6G).

Building of p62: The p62 protein was modeled based on the placement of the BSD domains (PDB ID 2DII), secondary structure prediction, extension of docked coordinates (Greber et al., 2017; Schilbach et al., 2017), and new tracing of the protein chain. Placement of the regions near XPD, where density is weak overall, was guided by matching the succession of secondary structure elements along the p62 sequence with helical densities in the cryo-EM map (Figure 2 – figure supplement 1A), and corroborated by CX-MS data (Luo et al., 2015), which showed excellent agreement of p62-XPD crosslinking data with the structure (Figure 1 – figure supplement 6G). Crosslinks between p62 and p44 showed a relatively large proportion of outliers (Supplementary Data Table 1), which may be due to the fact that the sequence register of the p62 segments to which these crosslinks map is not well constrained. These segments were modeled as poly-

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

alanines and deposited without sequence assignment (UNK; Supplementary Table 2). Maps low-pass filtered to 6 Å and sharpened by a b-factor of only -100 $Å^2$ were used to guide docking of domains and assess the continuity of the density in poorly ordered regions of the protein (Figure 2 – figure supplement 1B, C).

The resulting coordinate model (Supplementary Table 2) was refined using the real space refinement program in PHENIX (Adams et al., 2010; Afonine et al., 2018) and validated using the MTRIAGE program in PHENIX and the MOLPROBITY web server (Chen et al., 2010). Ramachandran, C_β, rotamer, and secondary structure restraints were used throughout the refinement to ensure good model geometry at the given resolution. Data used in the refinement excluded spatial frequencies beyond the nominal 3.7 Å resolution of the cryo-EM map to prevent over-fitting. The N-terminal RING domain of MAT1 and the BSD1 domain of p62, for which only poorly resolved density is present in the final cryo-EM map, were additionally restrained by reference restraints (Headd et al., 2012) using the NMR structures of the corresponding domains (PDB ID 1G25 and 2DII, respectively) (Gervais et al., 2001). The side chains of these two domains (with the exception of residues involved in zinc finger formation and of prolines) were truncated at the Cβ position to reflect the lower resolution of the corresponding densities. The FSC curve between the refined coordinate model and the cryo-EM map extends to 3.9 Å and the distribution of b-factors in the refined coordinate model (Figure 1 – figure supplement 2C) mirror the local resolution of the cryo-EM map (Figure 1 – figure supplement 2D), as expected. Refinement statistics are given in Supplementary Table 3 and are typical for structures in this resolution range (97th and 99th percentiles for MOLPROBITY clash score and overall score (Chen et al., 2010)).

Flexibility analysis

For the analysis of conformational dynamics of TFIIH, VPP datasets 1 and 2 were subjected to multibody refinement in RELION 3 (Nakane et al., 2018) using 6 masks (Figure 4 – figure supplement 1). The volume series for the first 16

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

principal components were reconstructed and difference densities (green and purple in Figure 4 – figure supplement 1) were computed between the most extreme states in each series (blue and yellow in Figure 4 – figure supplement 1). Subsequently, roughly 20,000 particles corresponding to both ends of the distribution were used for selected principal components and subjected to 3D refinement, resulting in maps of approx. 10 Å resolution. The refined atomic model of the TFIIH core complex, subdivided into suitable rigid bodies, was then rigid-body refined into these volumes using PHENIX real space refinement (Afonine et al., 2018) and coordinate displacement between the two resulting models for each principal component was plotted to obtain an initial assessment of the modes of motion present in the TFIIH dataset (Figure 4A). For actual structural interpretation (Figure 4B, C), the final cryo-EM maps of the TFIIH core complex (this work) and TFIIH in the context of the Pol II-PIC (Schilbach et al., 2017) were used.

Other

Figures were created using PyMol (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.) and the UCSF Chimera package from the Computer Graphics Laboratory, University of California, San Francisco (supported by NIH P41 RR-01081) (Pettersen et al., 2004). Protein-protein interface statistics were determined using PISA (Krissinel & Henrick, 2007).

ACKNOWLEDGMENTS

We thank D. Wigley for sharing of structural data, S. Scheres for sharing computer code before publication, R. Tjian, S. Zheng, and D. King for supplying XPB mAB and peptides for TFIIH purification, P. Grob for microscopy support, A. Chintangal and P. Tobias for computing support, and Z. Yu and C. Hong for data collection at the Janelia farm cryo-EM facility. We acknowledge the use of the resources of the National Energy Research Scientific Computing Center (NERSC), a DOE Office of Science user facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ADDITIONAL INFORMATION

Funding

This work was funded through NIGMS grants R01-GM63072 and P01-GM063210 to E.N.. B.J.G. was supported by fellowships from the Swiss National Science Foundation (projects P300PA_160983, P300PA_174355). E.N. is a Howard Hughes medical investigator.

Author contributions

E.N. directed the study. J.F. performed HeLa cell culture and purified TFIIH. B.J.G. performed cryo-EM specimen preparation, data processing, atomic model building, and coordinate refinement. D.B.T and B.J.G collected the cryo-EM data. B.J.G. wrote the initial draft of the manuscript and all authors contributed to the final version.

Supplementary Data Table 1 (separate file)

Result summary of mapping of CX-MS data (Luo et al., 2015) onto the TFIIH structure.

REFERENCES

Abdulrahman W., Iltis I., Radu L., Braun C., Maglott-Roth A., Giraudon C., Egly J.-M., Poterszman A. (2013) ARCH domain of XPD, an anchoring platform for CAK that conditions TFIIH DNA repair and transcription activities. *Proc Natl Acad Sci USA* **110:** E633-642

Adams P. D., Afonine P. V., Bunkóczi G., Chen V. B., Davis I. W., Echols N., Headd J. J., Hung L. W., Kapral G. J., Grosse-Kunstleve R. W. et al (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66:** 213-221

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Afonine P. V., Poon B. K., Read R. J., Sobolev O. V., Terwilliger T. C., Urzhumtsev A., Adams P. D. (2018) Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr D Struct Biol* **74:** 531-544

Alekseev S., Nagy Z., Sandoz J., Weiss A., Egly J.-M., Le May N., Coin F. (2017) Transcription without XPB Establishes a Unified Helicase-Independent Mechanism of Promoter Opening in Eukaryotic Gene Expression. *Mol Cell* **65:** 504-513.e505

Araújo S. J., Tirode F., Coin F., Pospiech H., Syväoja J. E., Stucki M., Hübscher U., Egly J. M., Wood R. D. (2000) Nucleotide excision repair of DNA with recombinant human proteins: definition of the minimal set of factors, active forms of TFIIH, and modulation by CAK. *Genes & Dev* **14:** 349-359

Bai X.-C., Rajendra E., Yang G., Shi Y., Scheres S. H. (2015) Sampling the conformational space of the catalytic subunit of human γ-secretase. *Elife* **4:** e11182

Balasingham S. V., Zegeye E. D., Homberset H., Rossi M. L., Laerdahl J. K., Bohr V. A., Tønjum T. (2012) Enzymatic activities and DNA substrate specificity of Mycobacterium tuberculosis DNA helicase XPB. *PLoS ONE* **7:** e36960

Berico P., Coin F. (2017) Is TFIIH the new Achilles heel of cancer cells? *Transcription* **9:** 47-51

Bienstock R. J., Skorvaga M., Mandavilli B. S., Van Houten B. (2003) Structural and functional characterization of the human DNA repair helicase XPD by comparative molecular modeling and site-directed mutagenesis of the bacterial repair protein UvrB. *J Biol Chem* **278:** 5309-5316

Biyani N., Righetto R. D., McLeod R., Caujolle-Bert D., Castaño-Diez D., Goldie K. N., Stahlberg H. (2017) Focus: The interface between data collection and data processing in cryo-EM. *J Struct Biol* **198:** 124-133

Botta E., Nardo T., Lehmann A. R., Egly J.-M., Pedrini A. M., Stefanini M. (2002) Reduced level of the repair/transcription factor TFIIH in trichothiodystrophy. *Hum Mol Genet* **11:** 2919-2928

Buechner C. N., Heil K., Michels G., Carell T., Kisker C., Tessmer I. (2014) Strand-specific recognition of DNA damages by XPD provides insights into nucleotide excision repair substrate versatility. *J Biol Chem* **289:** 3613-3624

Busso D., Keriel A., Sandrock B., Poterszman A., Gileadi O., Egly J. M. (2000) Distinct regions of MAT1 regulate cdk7 kinase and TFIIH transcription activities. *J Biol Chem* **275:** 22815-22823

Büttner K., Nehring S., Hopfner K.-P. (2007) Structural basis for DNA duplex separation by a superfamily-2 helicase. *Nat Struct Mol Biol* **14:** 647-652

Chen V. B., Arendall W. B., Headd J. J., Keedy D. A., Immormino R. M., Kapral G. J., Murray L. W., Richardson J. S., Richardson D. C. (2010) MolProbity: all-atom

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66:** 12-21

Cheng K., Wigley D. B. (2018) DNA translocation mechanism of an XPD family helicase. *Elife* **7**

Clarkson S. G., Wood R. D. (2005) Polymorphisms in the human XPD (ERCC2) gene, DNA repair capacity and cancer susceptibility: an appraisal. *DNA Repair (Amst)* **4:** 1068-1074

Cleaver J. E., Thompson L. H., Richardson A. S., States J. C. (1999) A summary of mutations in the UV-sensitive disorders: xeroderma pigmentosum, Cockayne syndrome, and trichothiodystrophy. *Hum Mutat* **14:** 9-22

Coin F., Marinoni J. C., Rodolfo C., Fribourg S., Pedrini A. M., Egly J. M. (1998) Mutations in the XPD helicase gene result in XP and TTD phenotypes, preventing interaction between XPD and the p44 subunit of TFIIH. *Nat Genet* **20:** 184-188

Coin F., Oksenych V., Egly J.-M. (2007) Distinct roles for the XPB/p52 and XPD/p44 subcomplexes of TFIIH in damaged DNA opening during nucleotide excision repair. *Mol Cell* **26:** 245-256

Coin F., Oksenych V., Mocquet V., Groh S., Blattner C., Egly J.-M. (2008) Nucleotide excision repair driven by the dissociation of CAK from TFIIH. *Mol Cell* **31:** 9-20

Coin F., Proietti De Santis L., Nardo T., Zlobinskaya O., Stefanini M., Egly J.-M. (2006) p8/TTD-A as a repair-specific TFIIH subunit. *Mol Cell* **21:** 215-226

Compe E., Egly J.-M. (2016) Nucleotide Excision Repair and Transcriptional Regulation: TFIIH and Beyond. *Annu Rev Biochem* **85:** 265-290

Constantinescu-Aruxandei D., Petrovic-Stojanovska B., Penedo J. C., White M. F., Naismith J. H. (2016) Mechanism of DNA loading by the DNA repair helicase XPD. *Nucleic Acids Res* **44:** 2806-2815

Danev R., Baumeister W. (2017) Expanding the boundaries of cryo-EM with phase plates. *Curr Opin Struct Biol* **46:** 87-94

Devault A., Martinez A. M., Fesquet D., Labbé J. C., Morin N., Tassan J. P., Nigg E. A., Cavadore J. C., Dorée M. (1995) MAT1 ('menage à trois') a new RING finger protein subunit stabilizing cyclin H-cdk7 complexes in starfish and Xenopus CAK. *EMBO J* **14:** 5027-5036

Di Lello P., Jenkins L. M. M., Jones T. N., Nguyen B. D., Hara T., Yamaguchi H., Dikeakos J. D., Appella E., Legault P., Omichinski J. G. (2006) Structure of the Tfb1/p53 complex: Insights into the interaction between the p62/Tfb1 subunit of TFIIH and the activation domain of p53. *Mol Cell* **22:** 731-740

Di Lello P., Miller Jenkins L. M., Mas C., Langlois C., Malitskaya E., Fradet-Turcotte A., Archambault J., Legault P., Omichinski J. G. (2008) p53 and TFIIEalpha share a

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

common binding site on the Tfb1/p62 subunit of TFIIH. *Proc Natl Acad Sci USA* **105:** 106-111

Drapkin R., Le Roy G., Cho H., Akoulitchev S., Reinberg D. (1996) Human cyclin-dependent kinase-activating kinase exists in three distinct complexes. *Proc Natl Acad Sci USA* **93:** 6488-6493

Dubaele S., Proietti De Santis L., Bienstock R. J., Keriel A., Stefanini M., Van Houten B., Egly J.-M. (2003) Basal transcription defect discriminates between xeroderma pigmentosum and trichothiodystrophy in XPD patients. *Mol Cell* **11:** 1635-1646

Emsley P., Lohkamp B., Scott W. G., Cowtan K. (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66:** 486-501

Evans E., Moggs J. G., Hwang J. R., Egly J. M., Wood R. D. (1997) Mechanism of open complex and dual incision formation by human nucleotide excision repair factors. *EMBO J* **16:** 6559-6573

Fagbemi A. F., Orelli B., Schärer O. D. (2011) Regulation of endonuclease activity in human nucleotide excision repair. *DNA Repair (Amst)* **10:** 722-729

Fairman-Williams M. E., Guenther U.-P., Jankowsky E. (2010) SF1 and SF2 helicases: family matters. *Curr Opin Struct Biol* **20:** 313-324

Fan L., Arvai A. S., Cooper P. K., Iwai S., Hanaoka F., Tainer J. A. (2006) Conserved XPB core structure and motifs for DNA unwinding: implications for pathway selection of transcription or excision repair. *Mol Cell* **22:** 27-37

Fan L., Fuss J. O., Cheng Q. J., Arvai A. S., Hammel M., Roberts V. A., Cooper P. K., Tainer J. A. (2008) XPD helicase structures and activities: insights into the cancer and aging phenotypes from XPD mutations. *Cell* **133:** 789-800

Fishburn J., Tomko E., Galburt E., Hahn S. (2015) Double-stranded DNA translocase activity of transcription factor TFIIH and the mechanism of RNA polymerase II open complex formation. *Proc Natl Acad Sci USA* **112:** 3961-3966

Fisher R. P., Jin P., Chamberlin H. M., Morgan D. O. (1995) Alternative mechanisms of CAK assembly require an assembly factor or an activating kinase. *Cell* **83:** 47-57

Fisher R. P., Morgan D. O. (1994) A novel cyclin associates with MO15/CDK7 to form the CDK-activating kinase. *Cell* **78:** 713-724

Fregoso M., Lainé J.-P., Aguilar-Fuentes J., Mocquet V., Reynaud E., Coin F., Egly J.-M., Zurita M. (2007) DNA repair and transcriptional deficiencies caused by mutations in the Drosophila p52 subunit of TFIIH generate developmental defects and chromosome fragility. *Mol Cell Biol* **27:** 3640-3650

Gervais V., Busso D., Wasielewski E., Poterszman A., Egly J. M., Thierry J. C., Kieffer B. (2001) Solution structure of the N-terminal domain of the human

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

TFIIH MAT1 subunit: new insights into the RING finger family. *J Biol Chem* **276:** 7457-7464

Gervais V., Lamour V., Jawhari A., Frindel F., Wasielewski E., Dubaele S., Egly J.-M., Thierry J.-C., Kieffer B., Poterszman A. (2004) TFIIH contains a PH domain involved in DNA nucleotide excision repair. *Nat Struct Mol Biol* **11:** 616-622

Gibbons B. J., Brignole E. J., Azubel M., Murakami K., Voss N. R., Bushnell D. A., Asturias F. J., Kornberg R. D. (2012) Subunit architecture of general transcription factor TFIIH. *Proc Natl Acad Sci USA* **109:** 1949-1954

Giglia-Mari G., Coin F., Ranish J. A., Hoogstraten D., Theil A., Wijgers N., Jaspers N. G. J., Raams A., Argentini M., van der Spek P. J. et al (2004) A new, tenth subunit of TFIIH is responsible for the DNA repair syndrome trichothiodystrophy group A. *Nat Genet* **36:** 714-719

Greber B. J., Nguyen T. H. D., Fang J., Afonine P. V., Adams P. D., Nogales E. (2017) The cryo-electron microscopy structure of human transcription factor IIH. *Nature* **549:** 414-417

Grünberg S., Warfield L., Hahn S. (2012) Architecture of the RNA polymerase II preinitiation complex and mechanism of ATP-dependent promoter opening. *Nat Struct Mol Biol* **19:** 788-796

Guzder S. N., Qiu H., Sommers C. H., Sung P., Prakash L., Prakash S. (1994) DNA repair gene RAD3 of S. cerevisiae is essential for transcription by RNA polymerase II. *Nature* **367:** 91-94

He Y., Fang J., Taatjes D. J., Nogales E. (2013) Structural visualization of key steps in human transcription initiation. *Nature* **495:** 481-486

He Y., Yan C., Fang J., Inouye C., Tjian R., Ivanov I., Nogales E. (2016) Near-atomic resolution visualization of human transcription promoter opening. *Nature* **533:** 359-365

Headd J. J., Echols N., Afonine P. V., Grosse-Kunstleve R. W., Chen V. B., Moriarty N. W., Richardson D. C., Richardson J. S., Adams P. D. (2012) Use of knowledge-based restraints in phenix.refine to improve macromolecular refinement at low resolution. *Acta Crystallogr D Biol Crystallogr* **68:** 381-390

Herrera-Moyano E., Moriel-Carretero M., Montelone B. A., Aguilera A. (2014) The rem mutations in the ATP-binding groove of the Rad3/XPD helicase lead to Xeroderma pigmentosum-Cockayne syndrome-like phenotypes. *PLoS Genet* **10:** e1004859

Hilario E., Li Y., Nobumori Y., Liu X., Fan L. (2013) Structure of the C-terminal half of human XPB helicase and the impact of the disease-causing mutation XP11BE. *Acta Crystallogr D Biol Crystallogr* **69:** 237-246

Huang J. C., Svoboda D. L., Reardon J. T., Sancar A. (1992) Human nucleotide excision nuclease removes thymine dimers from DNA by incising the 22nd

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

phosphodiester bond 5' and the 6th phosphodiester bond 3' to the photodimer. *Proc Natl Acad Sci USA* **89:** 3664-3668

Jawhari A., Boussert S., Lamour V., Atkinson R. A., Kieffer B., Poch O., Potier N., Van Dorsselaer A., Moras D., Poterszman A. (2004) Domain architecture of the p62 subunit from the human transcription/repair factor TFIIH deduced by limited proteolysis and mass spectrometry analysis. *Biochemistry* **43:** 14420-14430

Jawhari A., Lainé J.-P., Dubaele S., Lamour V., Poterszman A., Coin F., Moras D., Egly J.-M. (2002) p52 Mediates XPB function within the transcription/repair factor TFIIH. *J Biol Chem* **277:** 31761-31767

Jones T. A., Zou J. Y., Cowan S. W., Kjeldgaard M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A Found Crystallogr* **47 (Pt. 2):** 110-119

Kainov D. E., Vitorino M., Cavarelli J., Poterszman A., Egly J.-M. (2008) Structural basis for group A trichothiodystrophy. *Nat Struct Mol Biol* **15:** 980-984

Kelley L. A., Mezulis S., Yates C. M., Wass M. N., Sternberg M. J. E. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10:** 845-858

Khoshouei M., Radjainia M., Baumeister W., Danev R. (2017) Cryo-EM structure of haemoglobin at 3.2 Å determined with the Volta phase plate. *Nat Commun* **8:** 16099

Kim J. S., Saint-André C., Lim H. S., Hwang C.-S., Egly J.-M., Cho Y. (2015) Crystal structure of the Rad3/XPD regulatory domain of Ssl1/p44. *J Biol Chem* **290:** 8321-8330

Kimanius D., Forsberg B. O., Scheres S. H., Lindahl E. (2016) Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. *Elife* **5:** e18722

Kralund H. H., Ousager L., Jaspers N. G., Raams A., Pedersen E. B., Gade E., Bygum A. (2013) Xeroderma Pigmentosum-Trichothiodystrophy overlap patient with novel XPD/ERCC2 mutation. *Rare Dis* **1:** e24932

Krissinel E., Henrick K. (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372:** 774-797

Kuper J., Braun C., Elias A., Michels G., Sauer F., Schmitt D. R., Poterszman A., Egly J.-M., Kisker C. (2014) In TFIIH, XPD helicase is exclusively devoted to DNA repair. *PLoS Biol* **12:** e1001954

Kuper J., Wolski S. C., Michels G., Kisker C. (2012) Functional and structural studies of the nucleotide excision repair helicase XPD suggest a polarity for DNA translocation. *EMBO J* **31:** 494-502

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Lafrance-Vanasse J., Arseneault G., Cappadocia L., Legault P., Omichinski J. G. (2013) Structural and functional evidence that Rad4 competes with Rad2 for binding to the Tfb1 subunit of TFIIH in NER. *Nucleic Acids Res* **41:** 2736-2745

Li C.-L., Golebiowski F. M., Onishi Y., Samara N. L., Sugasawa K., Yang W. (2015) Tripartite DNA Lesion Recognition and Verification by XPC, TFIIH, and XPA in Nucleotide Excision Repair. *Mol Cell* **59:** 1025-1034

Liu H., Rudolf J., Johnson K. A., McMahon S. A., Oke M., Carter L., McRobbie A.-M., Brown S. E., Naismith J. H., White M. F. (2008) Structure of the DNA Repair Helicase XPD. *Cell* **133:** 801-812

Liu J., Huang C., Shin D.-H., Yokota H., Jancarik J., Kim J.-S., Adams P. D., Kim R., Kim S.-H. (2005) Crystal structure of a heat-inducible transcriptional repressor HrcA from Thermotoga maritima: structural insight into DNA binding and dimerization. *J Mol Biol* **350:** 987-996

Luo J., Cimermančič P., Viswanath S., Ebmeier C. C., Kim B., Dehecq M., Raman V., Greenberg C. H., Pellarin R., Sali A. et al (2015) Architecture of the Human and Yeast General Transcription and DNA Repair Factor TFIIH. *Mol Cell* **59:** 794-806

Mason A. C., Rambo R. P., Greer B., Pritchett M., Tainer J. A., Cortez D., Eichman B. F. (2014) A structure-specific nucleic acid-binding domain conserved among DNA repair proteins. *Proc Natl Acad Sci USA* **111:** 7618-7623

Mathieu N., Kaczmarek N., Rüthemann P., Luch A., Naegeli H. (2013) DNA quality control by a lesion sensor pocket of the xeroderma pigmentosum group D helicase subunit of TFIIH. *Curr Biol* **23:** 204-212

Mlak R., Krawczyk P., Homa-Mlak I., Powrózek T., Ciesielka M., Kozioł P., Milanowski J., Małecka-Massalska T. (2018) Predictive Value of Single Nucleotide Polymorphisms of ERCC1, XPA, XPC, XPD and XPG Genes, Involved in NER Mechanism in Patients with Advanced NSCLC Treated with Cisplatin and Gemcitabine. *Pathol Oncol Res*: 1-11

Moriel-Carretero M., Aguilera A. (2010) A postincision-deficient TFIIH causes replication fork breakage and uncovers alternative Rad51- or Pol32-mediated restart mechanisms. *Mol Cell* **37:** 690-701

Moriel-Carretero M., Herrera-Moyano E., Aguilera A. (2015) A unified model for the molecular basis of Xeroderma pigmentosum-Cockayne Syndrome. *Rare Dis* **3:** e1079362

Murakami K., Tsai K.-L., Kalisman N., Bushnell D. A., Asturias F. J., Kornberg R. D. (2015) Structure of an RNA polymerase II preinitiation complex. *Proc Natl Acad Sci USA* **112:** 13543-13548

Naegeli H., Bardwell L., Friedberg E. C. (1993) Inhibition of Rad3 DNA helicase activity by DNA adducts and abasic sites: implications for the role of a DNA helicase in damage-specific incision of DNA. *Biochemistry* **32:** 613-621

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Nakane T., Kimanius D., Lindahl E., Scheres S. H. (2018) Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *Elife* **7:** e36861

Nguyen T. H. D., Galej W. P., Bai X.-C., Oubridge C., Newman A. J., Scheres S. H. W., Nagai K. (2016) Cryo-EM structure of the yeast U4/U6.U5 tri-snRNP at 3.7 Å resolution. *Nature* **530:** 298-302

Nguyen T. H. D., Galej W. P., Bai X.-C., Savva C. G., Newman A. J., Scheres S. H. W., Nagai K. (2015) The architecture of the spliceosomal U4/U6.U5 tri-snRNP. *Nature* **523:** 47-52

Obmolova G., Ban C., Hsieh P., Yang W. (2000) Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA. *Nature* **407:** 703-710

Oksenych V., Bernardes de Jesus B., Zhovmer A., Egly J.-M., Coin F. (2009) Molecular insights into the recruitment of TFIIH to sites of DNA damage. *EMBO J* **28:** 2971-2980

Okuda M., Nakazawa Y., Guo C., Ogi T., Nishimura Y. (2017) Common TFIIH recruitment mechanism in global genome and transcription-coupled repair subpathways. *Nucleic Acids Res* **45:** 13043-13055

Peters G. J., Avan A., Ruiz M. G., Orsini V., Avan A., Giovannetti E., Smit E. F. (2014) Predictive role of repair enzymes in the efficacy of Cisplatin combinations in pancreatic and lung cancer. *Anticancer Res* **34:** 435-442

Pettersen E. F., Goddard T. D., Huang C. C., Couch G. S., Greenblatt D. M., Meng E. C., Ferrin T. E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25:** 1605-1612

Poterszman A., Lamour V., Egly J. M., Moras D., Thierry J. C., Poch O. (1997) A eukaryotic XPB/ERCC3-like helicase in Mycobacterium leprae? *Trends Biochem Sci* **22:** 418-419

Radu L., Schoenwetter E., Braun C., Marcoux J., Koelmel W., Schmitt D. R., Kuper J., Cianférani S., Egly J. M., Poterszman A. et al (2017) The intricate network between the p34 and p44 subunits is central to the activity of the transcription/DNA repair factor TFIIH. *Nucleic Acids Res* **45:** 10872-10883

Rapin I. (2013) Disorders of nucleotide excision repair. *Handb Clin Neurol* **113:** 1637-1650

Riou L., Zeng L., Chevallier-Lagente O., Stary A., Nikaido O., Taïeb A., Weeda G., Mezzina M., Sarasin A. (1999) The relative expression of mutated XPB genes results in xeroderma pigmentosum/Cockayne's syndrome or trichothiodystrophy cellular phenotypes. *Hum Mol Genet* **8:** 1125-1133

Robert X., Gouet P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* **42:** W320-324

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Robinson P. J., Trnka M. J., Bushnell D. A., Davis R. E., Mattei P.-J., Burlingame A. L., Kornberg R. D. (2016) Structure of a Complete Mediator-RNA Polymerase II Pre-Initiation Complex. *Cell* **166:** 1411-1422.e1416

Rosenthal P. B., Henderson R. (2003) Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J Mol Biol* **333:** 721-745

Rossignol M., Kolb-Cheynel I., Egly J. M. (1997) Substrate specificity of the cdk-activating kinase (CAK) is altered upon association with TFIIH. *EMBO J* **16:** 1628-1637

Rouillon C., White M. F. (2010) The XBP-Bax1 helicase-nuclease complex unwinds and cleaves DNA: implications for eukaryal and archaeal nucleotide excision repair. *J Biol Chem* **285:** 11013-11022

Rudolf J., Makrantoni V., Ingledew W. J., Stark M. J. R., White M. F. (2006) The DNA repair helicases XPD and FancJ have essential iron-sulfur domains. *Mol Cell* **23:** 801-808

Sainsbury S., Bernecky C., Cramer P. (2015) Structural basis of transcription initiation by RNA polymerase II. *Nat Rev Mol Cell Biol* **16:** 129-143

Sandrock B., Egly J. M. (2001) A yeast four-hybrid system identifies Cdk-activating kinase as a regulator of the XPD helicase, a subunit of transcription factor IIH. *J Biol Chem* **276:** 35328-35333

Scheres S. H. W. (2010) Classification of Structural Heterogeneity by Maximum-Likelihood Methods. *Meth Enz* **482:** 295-320

Scheres S. H. W. (2012) RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* **180:** 519-530

Scheres S. H. W. (2015) Semi-automated selection of cryo-EM particles in RELION-1.3. *J Struct Biol* **189:** 114-122

Scheres S. H. W., Chen S. (2012) Prevention of overfitting in cryo-EM structure determination. *Nat Meth* **9:** 853-854

Schilbach S., Hantsche M., Tegunov D., Dienemann C., Wigge C., Urlaub H., Cramer P. (2017) Structures of transcription pre-initiation complex with TFIIH and Mediator. *Nature* **551:** 204-209

Seroz T., Perez C., Bergmann E., Bradsher J., Egly J. M. (2000) p44/SSL1, the regulatory subunit of the XPD/RAD3 helicase, plays a crucial role in the transcriptional activity of TFIIH. *J Biol Chem* **275:** 33260-33266

Shiekhattar R., Mermelstein F., Fisher R. P., Drapkin R., Dynlacht B., Wessling H. C., Morgan D. O., Reinberg D. (1995) Cdk-activating kinase complex is a component of human transcription factor TFIIH. *Nature* **374:** 283-287

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Sugasawa K., Akagi J.-i., Nishi R., Iwai S., Hanaoka F. (2009) 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

Sugitani N., Sivley R. M., Perry K. E., Capra J. A., Chazin W. J. (2016) XPA: A key scaffold for human nucleotide excision repair. *DNA Repair (Amst)* **44:** 123-135

Svejstrup J. Q., Wang Z., Feaver W. J., Wu X., Bushnell D. A., Donahue T. F., Friedberg E. C., Kornberg R. D. (1995) Different forms of TFIIH for transcription and DNA repair: holo-TFIIH and a nucleotide excision repairosome. *Cell* **80:** 21-28

Theron T., Fousteri M. I., Volker M., Harries L. W., Botta E., Stefanini M., Fujimoto M., Andressoo J.-O., Mitchell J., Jaspers N. G. J. et al (2005) Transcription-associated breaks in xeroderma pigmentosum group D cells from patients with combined features of xeroderma pigmentosum and Cockayne syndrome. *Mol Cell Biol* **25:** 8368-8378

Tirode F., Busso D., Coin F., Egly J. M. (1999) Reconstitution of the transcription factor TFIIH: assignment of functions for the three enzymatic subunits, XPB, XPD, and cdk7. *Mol Cell* **3:** 87-95

Tremeau-Bravard A., Perez C., Egly J. M. (2001) A role of the C-terminal part of p44 in the promoter escape activity of transcription factor IIH. *J Biol Chem* **276:** 27693-27697

Tsai K.-L., Yu X., Gopalan S., Chao T.-C., Zhang Y., Florens L., Washburn M. P., Murakami K., Conaway R. C., Conaway J. W. et al (2017) Mediator structure and rearrangements required for holoenzyme formation. *Nature Publishing Group*: 1-19

Vermeulen W., Rademakers S., Jaspers N. G., Appeldoorn E., Raams A., Klein B., Kleijer W. J., Hansen L. K., Hoeijmakers J. H. (2001) A temperature-sensitive disorder in basal transcription and DNA repair in humans. *Nat Genet* **27:** 299-303

Vitorino M., Coin F., Zlobinskaya O., Atkinson R. A., Moras D., Egly J.-M., Poterszman A., Kieffer B. (2007) Solution structure and self-association properties of the p8 TFIIH subunit responsible for trichothiodystrophy. *J Mol Biol* **368:** 473-480

Warfield L., Luo J., Ranish J., Hahn S. (2016) Function of conserved topological regions within the S. cerevisiae basal transcription factor TFIIH. *Mol Cell Biol* **36:** 2464-2475

Winkler G. S., Araújo S. J., Fiedler U., Vermeulen W., Coin F., Egly J. M., Hoeijmakers J. H., Wood R. D., Timmers H. T., Weeda G. (2000) TFIIH with inactive XPD helicase functions in transcription initiation but is defective in DNA repair. *J Biol Chem* **275:** 4258-4266

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Wirth N., Gross J., Roth H. M., Buechner C. N., Kisker C., Tessmer I. (2016) Conservation and Divergence in Nucleotide Excision Repair Lesion Recognition. *J Biol Chem* **291:** 18932-18946

Wolski S. C., Kuper J., Hänzelmann P., Truglio J. J., Croteau D. L., Van Houten B., Kisker C. (2008) Crystal structure of the FeS cluster-containing nucleotide excision repair helicase XPD. *PLoS Biol* **6:** e149

Zhang K. (2016) Gctf: Real-time CTF determination and correction. *J Struct Biol* **193:** 1-12

Zheng S. Q., Palovcak E., Armache J.-P., Verba K. A., Cheng Y., Agard D. A. (2017) MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Meth* **14:** 331-332

Ziani S., Nagy Z., Alekseev S., Soutoglou E., Egly J.-M., Coin F. (2014) Sequential and ordered assembly of a large DNA repair complex on undamaged chromatin. *J Cell Biol* **206:** 589-598

Zivanov J., Nakane T., Forsberg B. O., Kimanius D., Hagen W. J., Lindahl E., Scheres S. H. (2018) New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7:** e42166

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure supplements

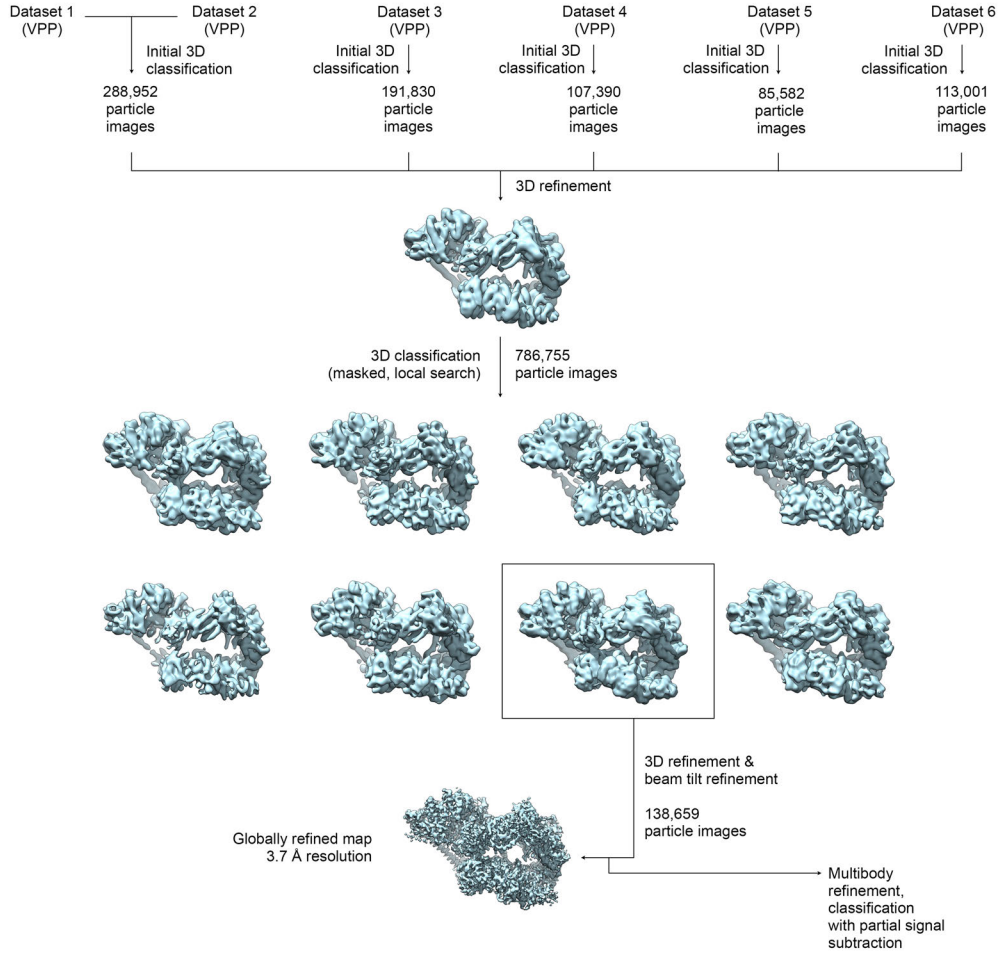


Figure 1 - figure supplement 1. Data processing scheme for global reconstruction. The datasets (DS) were initially processed individually, with the exception of DS6, where particles from DS4 were used to compensate for orientation bias, and DS1/DS2, which were joined for processing. Subsequently, the datasets were pooled and sub-classified. The best class of was used for the final refinement to obtain a globally refined map at 3.7 Å resolution (see Figure 1 – figure supplement 2). The classified VPP dataset was also used for further processing and classification of sub-regions (Figure 1 – figure supplements 3-5).

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

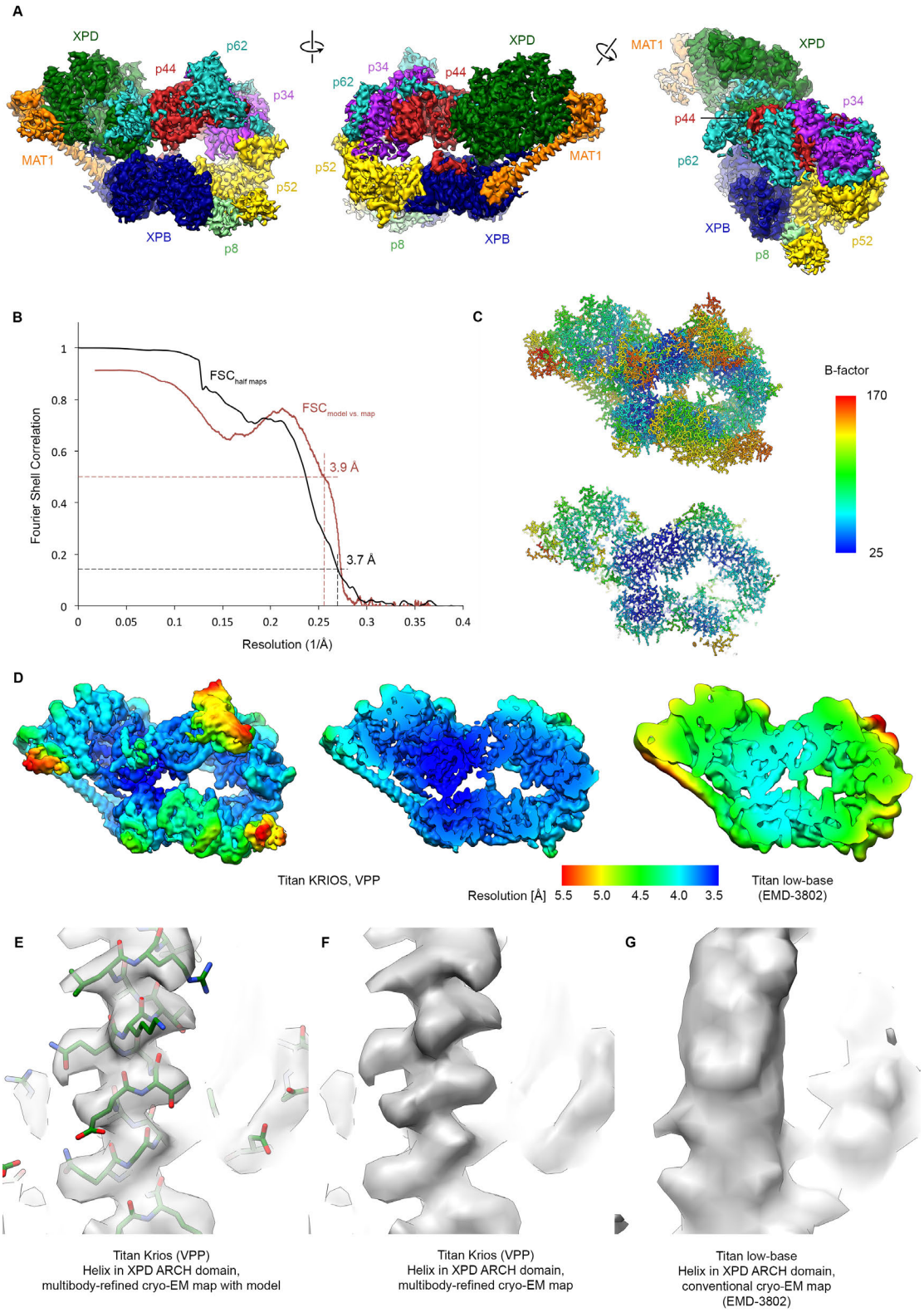


Figure 1 – figure supplement 2. Resolution estimation, validation statistics, and quality of the density. (**A**) Three views of the refined, sharpened, and low-pass filtered cryo-EM map of the TFIIH core complex. Protein subunits are color-coded and labeled. (**B**) Fourier shell correlation (FSC) curves. Black: FSC curve

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

between half-maps. Red: FSC curve between map and refined model. The resolution at the appropriate thresholds (Rosenthal & Henderson, 2003; Scheres & Chen, 2012) is indicated. (**C**) B-factor distribution on the refined TFIIH coordinate model. (**D**) Local resolution estimation, computed by windowed FSC in RELION 2. (**E**, **F**, **G**) Comparison of the 3.6 Å multibody-refined cryo-EM map of the TFIIH core complex (E, F) with our previous 4.4 Å cryo-EM map (G) shows substantial improvements in map quality in the more peripheral and flexible regions of the complex (an α-helix from the XPD ARCH domain is shown).

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

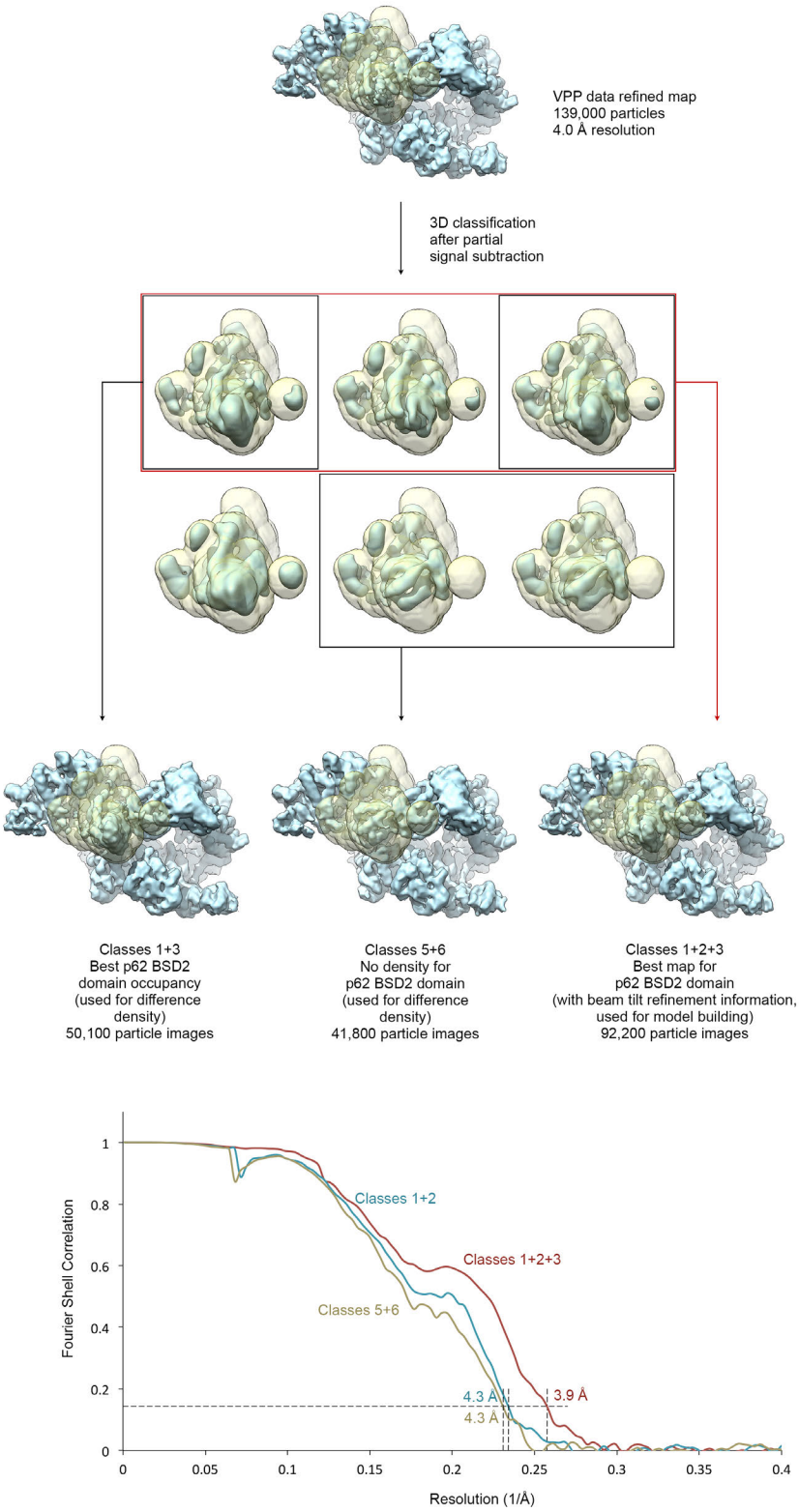


Figure 1 – figure supplement 3. Focused classification of the p62 BSD2 domain. (**A**) Focused classification for the p62 BSD2 domain and adjacent structural elements using the classified VPP dataset (indicated in Figure 1 –

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

figure supplement 1). Partial signal subtraction (Bai et al., 2015) and masking were used to focus the classification on the region of interest (mask shown in transparent yellow). Selected classes were refined using the selected but un-subtracted particle images because the region of interest is too small for efficient image alignment (Nguyen et al., 2016). (**B**) FSC curves for the three refined maps shown in (A), indicating resolution of 3.9-4.3 Å according to the FSC = 0.143 criterion (Rosenthal & Henderson, 2003).

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

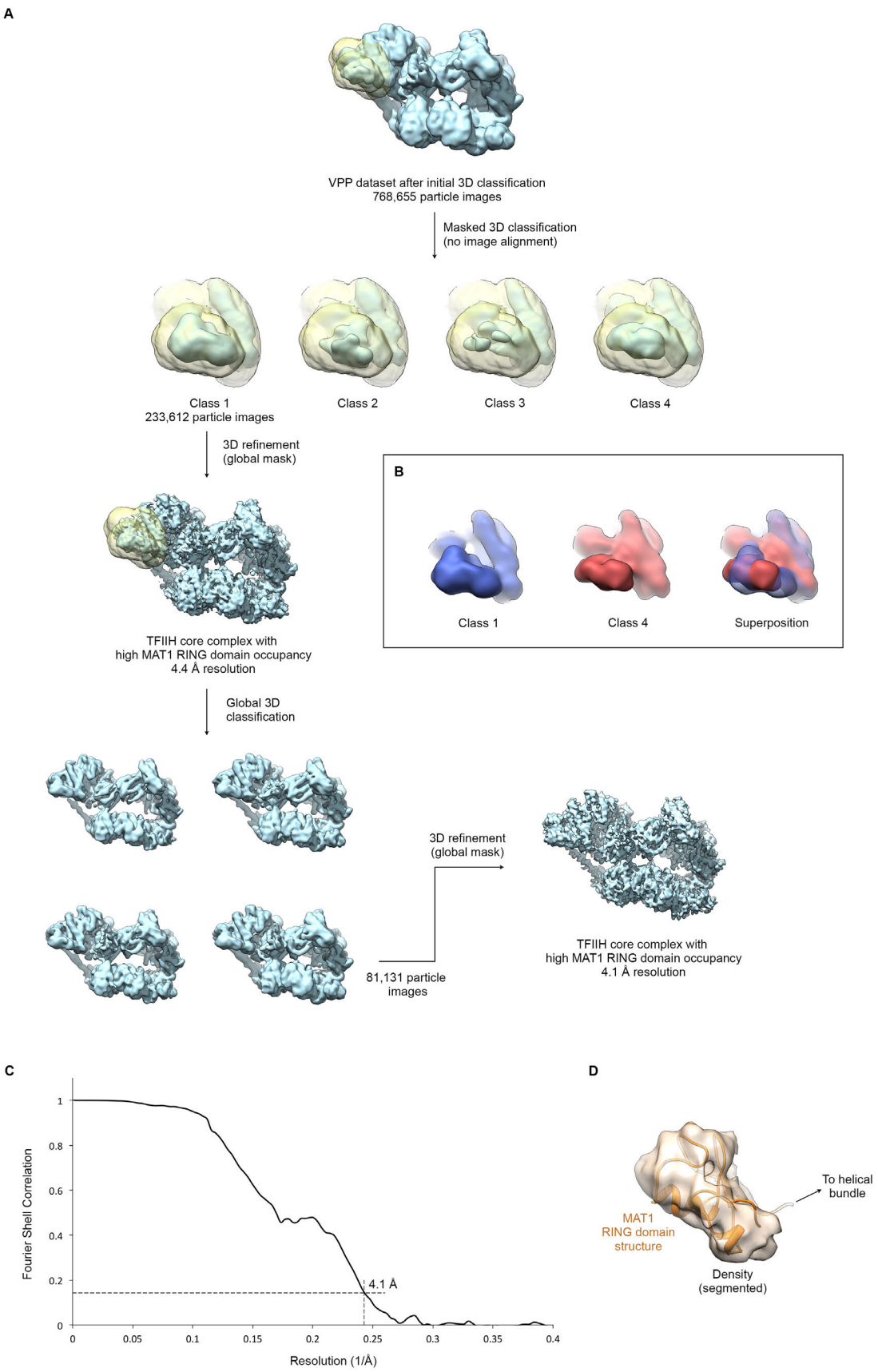


Figure 1 – figure supplement 4. Focused classification and interpretation of the MAT1 RING domain density. (**A**) The entire pool of pre-classified TFIIH

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

particles from the VPP datasets was subjected to alignment-free focused 3D classification using a mask that covers the MAT1 RING domain and helical bundle. Signal subtraction was not required due to the peripheral location of the region in the complex and the orientation distribution of the particle images. After one round of additional global classification (to select for overall particle quality and conformation), the class showing the most highly occupied density for the MAT1 RING domain was refined to 4.1 Å resolution (using beam tilt values imported from the best-resolution refinement, Figure 1 – figure supplement 1). (**B**) 3D classes differed not only in occupancy (i.e. classes 2 and 3 showed low occupancy and fragmented density for the MAT1 RING domain) but also in the orientation of the MAT1 RING domain, which is flexibly attached to the complex. Classes 1 and 4 are shown in different colors individually and superposed to highlight the different orientations of the MAT1 RING domain. (**C**) The FSC curve estimates a resolution of 4.1 Å for the refined map of TFIIH with high occupancy of the MAT1 RING domain. (**D**) Best fit of the structure of the MAT1 RING domain (Gervais et al., 2001) into the cryo-EM map, obtained using the FITMAP command in UCSF CHIMERA (Pettersen et al., 2004) with 100 independent initial placements before local optimization at 5 Å resolution. This fit shows the best correlation between model and map and additionally satisfies geometrical constraints imposed by the continuity of the peptide chain between the RING domain and the MAT1 helical bundle.

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

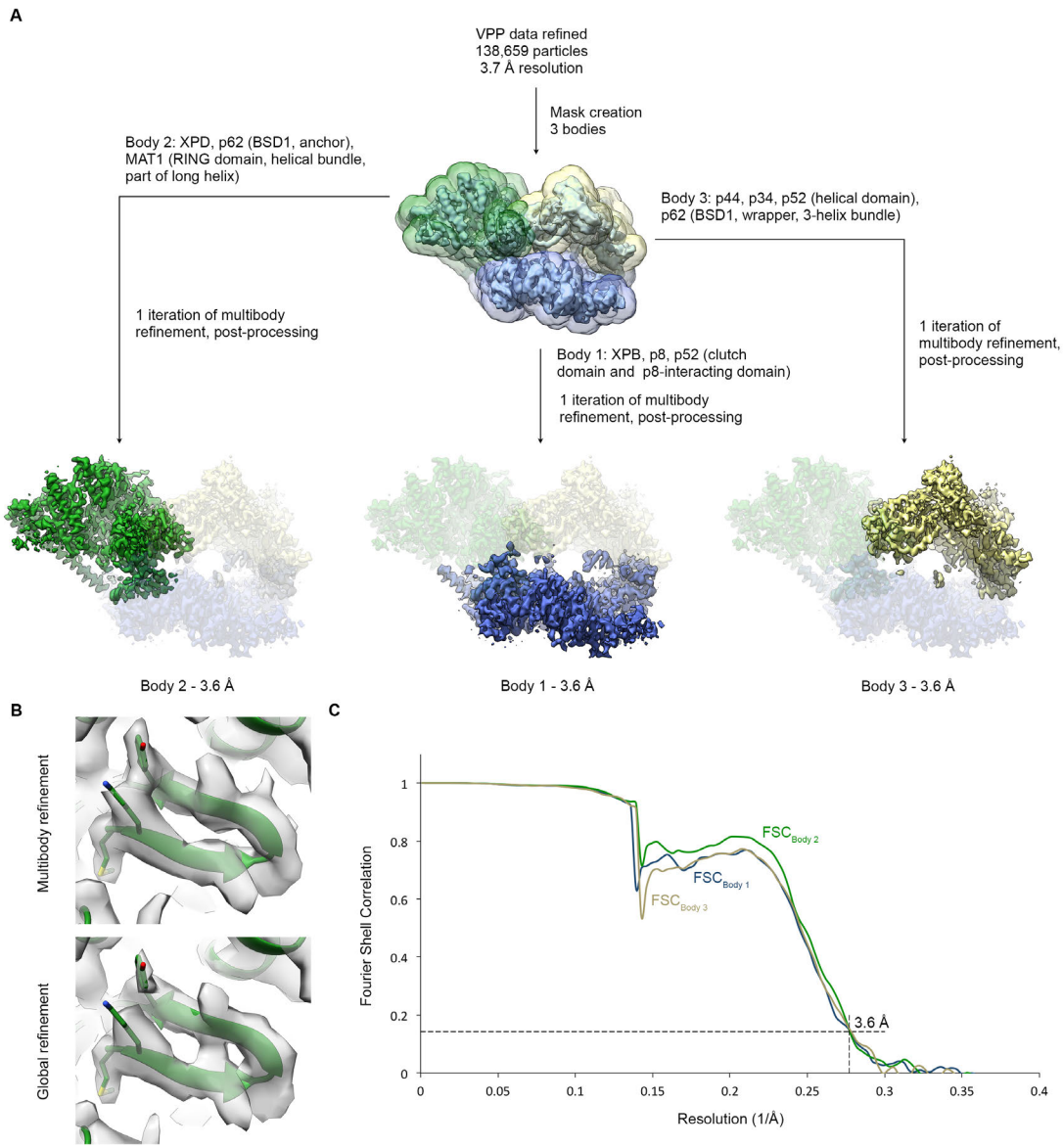


Figure 1 – figure supplement 5. Multibody refinement. (**A**) Schematic for multibody refinement (Nakane et al., 2018) of the classified VPP dataset (indicated in Figure 1 – figure supplement 1). Three bodies were used and subjected to multibody refinement using the corresponding masks. (**B**) Example of the improvement of the cryo-EM density for the XPD N-terminus after multibody refinement (top) relative to the global refinement (bottom). (**C**) FSC curves for the refined bodies, indicating resolutions of 3.6 Å according to the FSC = 0.143 criterion (Rosenthal & Henderson, 2003).

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

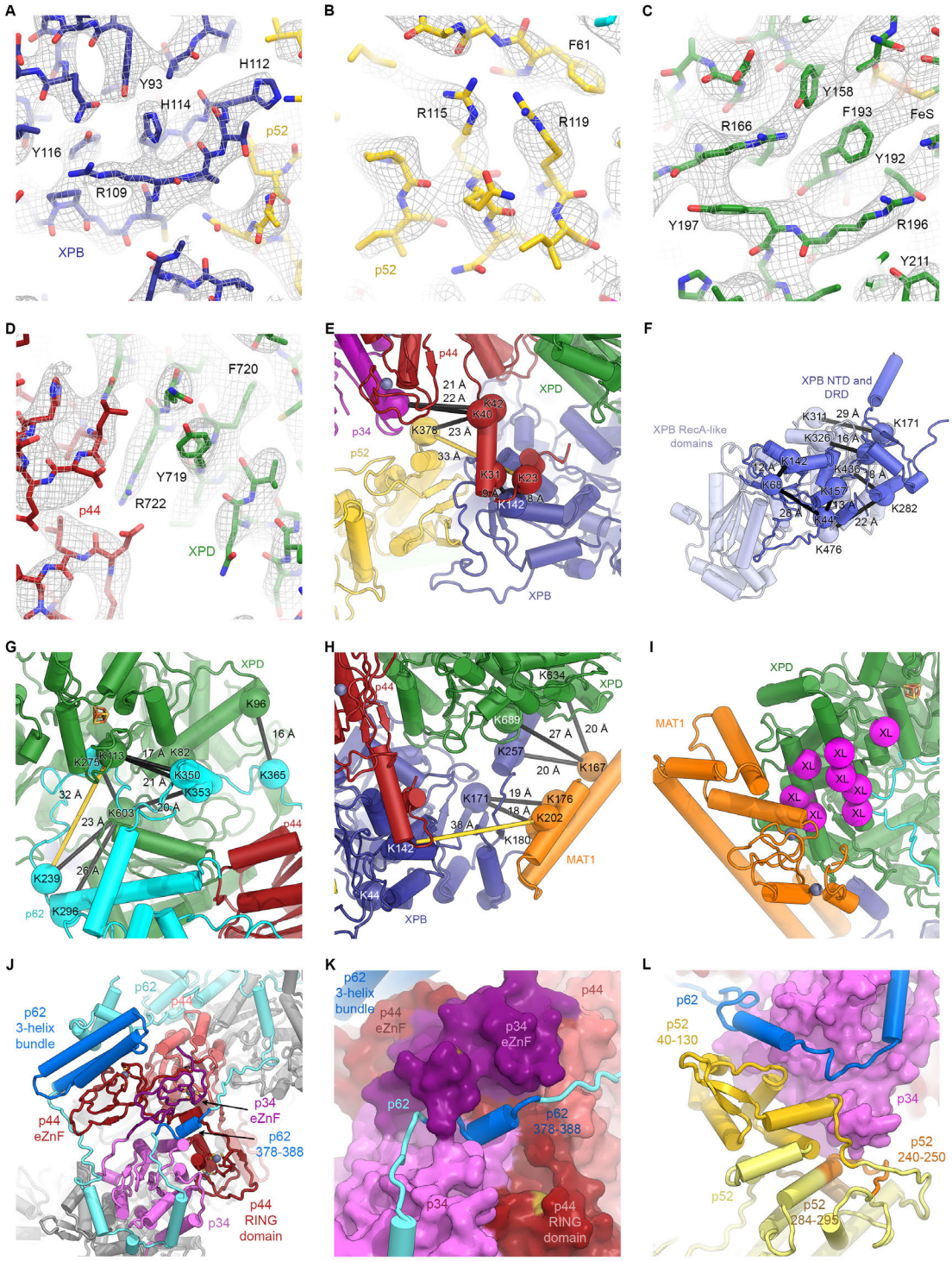


Figure 1 – figure supplement 6. Detailed architecture of the TFIIH core complex. (**A-D**) The refined coordinate model is shown in the 3.7 Å-resolution cryo-EM map. Large side chains are indicated. (**E-H**) Mapping of CX-MS data (Luo et al., 2015) onto the atomic coordinate model. Crosslinked residues are shown as spheres, crosslinks as black lines (crosslinks > 30 Å apart as yellow lines). (**E**) Crosslinks between the p44 NTE and XPB, p52, and p34 confirm the tracing of

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the p44 NTE towards the XPB NTD. (**F**) Intramolecular crosslinks in XPB validate the tracing of the XPB NTD. (**G**) Crosslinks between p62 and XPD. Crosslinks between p62 and p44 are not shown (see Methods and Supplementary Data Table 1). (**H**) Crosslinks between MAT1 and XPB and XPD. One strong outlier connecting XPB K44 to the MAT1 helical bundle (>100 Å distance) is a likely false positive and is not shown. (**I**) Location of site-specific crosslinks between RAD3, the yeast XPD homolog (RAD3 residues carrying the crosslinker mapped onto the XPD structure and shown as purple spheres) and TFB3, the yeast MAT1 homolog (Warfield et al., 2016). (**J**) Interlocking zinc-binding domains of p44 and p34 mediate interactions between these proteins in the hinge region. The p62 C-terminal 3-helix bundle binds to the p44 eZnF domain. (**K**) Residues 378-388 of p62 are sandwiched between the zinc-binding domains of p34 and p44. (**L**) View of the p62-p52-p34 interaction region.

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

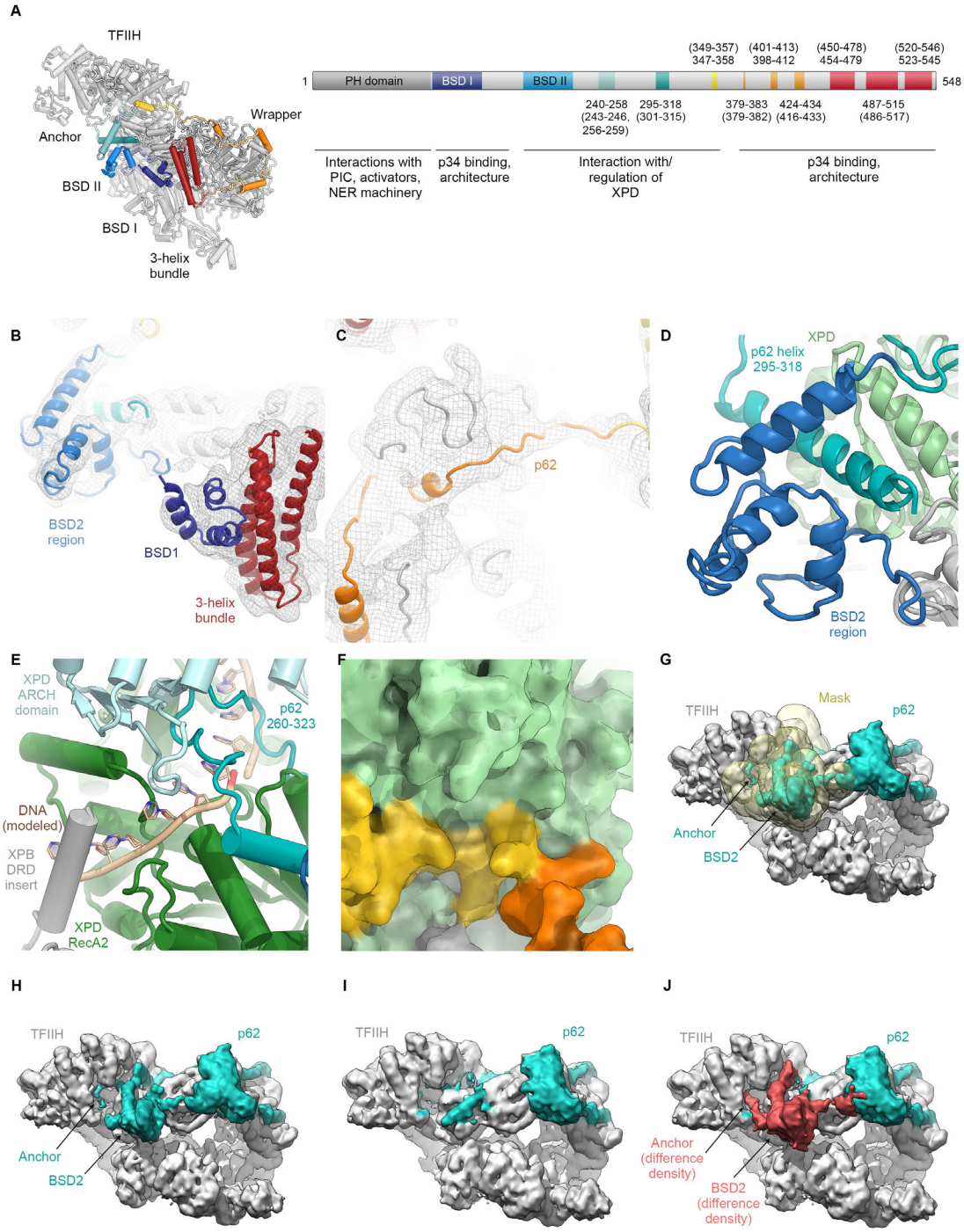


Figure 2 – figure supplement 1. Analysis of the structure of p62. (**A**) p62 architecture and sequence assignment. Fitting of domains (see **B**), secondary structure prediction for α-helices (Kelley et al., 2015), helical segments of the cryo-EM map, and density for large side chains were used for initial assignment of the trace of p62. The final model shows good agreement of observed helical segments (residue numbers indicated) and secondary structure prediction for α-

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

helices (residue numbers given in brackets). (**B**) Fitting of p62 BSD1 (PDB ID 2DII; blue) into the TFIIH cryo-EM map. Due to flexibility, the density in this region is weak and was low-pass filtered to 6 Å for interpretation and fitting. In agreement with previous results (Schilbach et al., 2017), this domain interacts with the p62 C-terminal three-helix bundle (purple), which in turn sits on the eZnF domain of p44 (Figure 1 – figure supplement 6J), thereby explaining the importance of the latter region for functional TFIIH (Tremeau-Bravard et al., 2001). (**C**) Extended segments of the p62 chain were interpreted using the low-pass filtered map as described in (B) to maintain continuity of the cryo-EM density. (**D**) The p62 helix 295-317 is sandwiched between the BSD2 region of p62 and XPD RecA2, thereby bridging these two elements of TFIIH. (**E**) XPD-bound DNA was modeled by domain-wise superposition of the structure of the substrate-bound helicase DinG (Cheng & Wigley, 2018) onto the RecA-like domains of XPD. The presence of the p62 anchor near XPD RecA2 leads to steric hindrance with the DNA that would prevent binding. (**F**) Residues 346-361 of p62 block access to the XPD nucleotide-binding pocket. (**G-J**) Analysis of p62 (teal) conformational dynamics. (**G**) Particle images were sorted for p62 elements near XPD using signal-subtracted focused classification (mask shown in yellow; see Figure 1 – figure supplement 3). (**H**) Refined class with p62 density near XPD. (**I**) Refined class without p62 density near XPD. (**J**) Difference density (red) between the densities shown in H, I. Both classes show intact particles, even in the absence of the p62 segment inserted into the XPD substrate binding cavity.

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

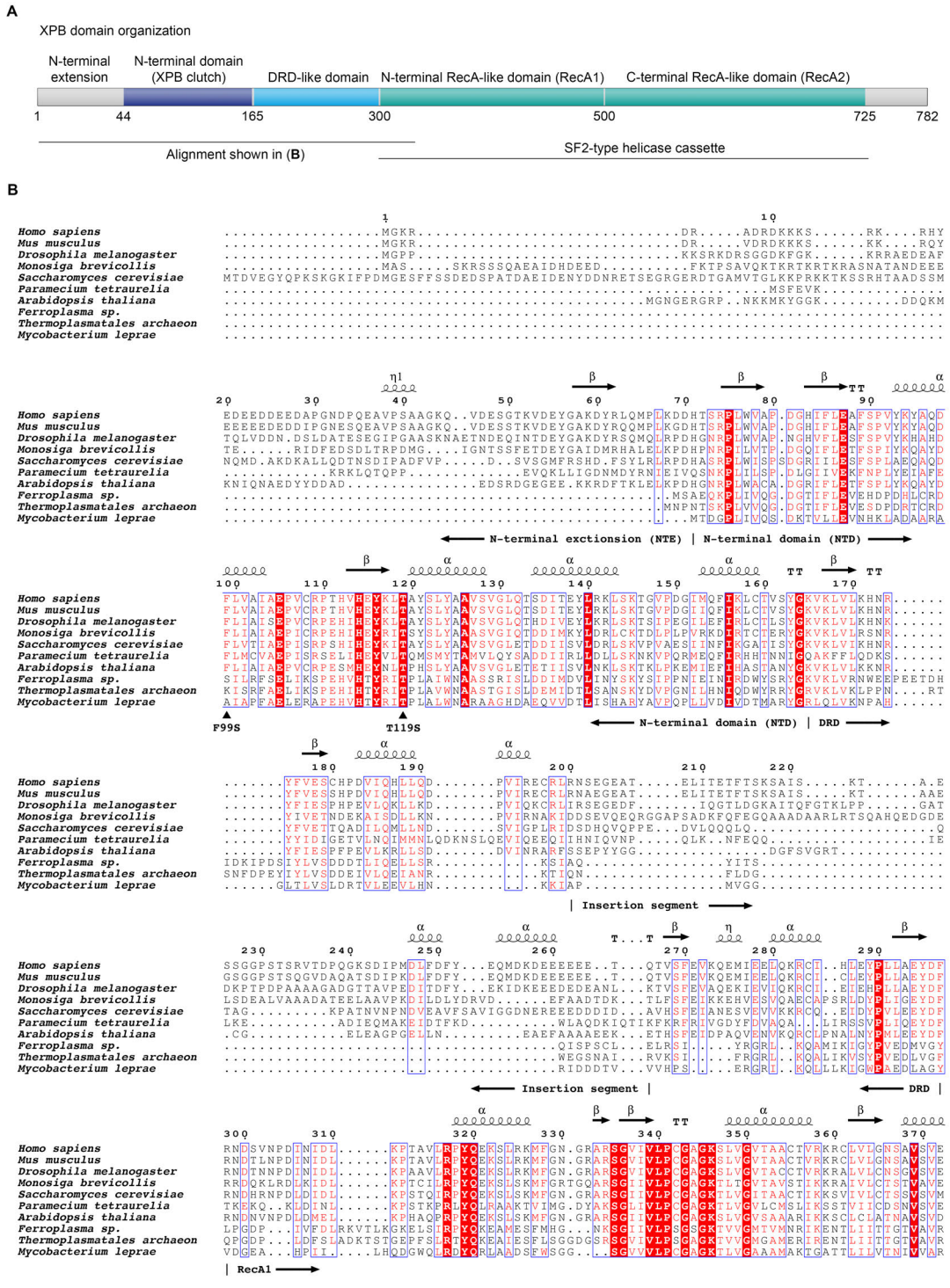
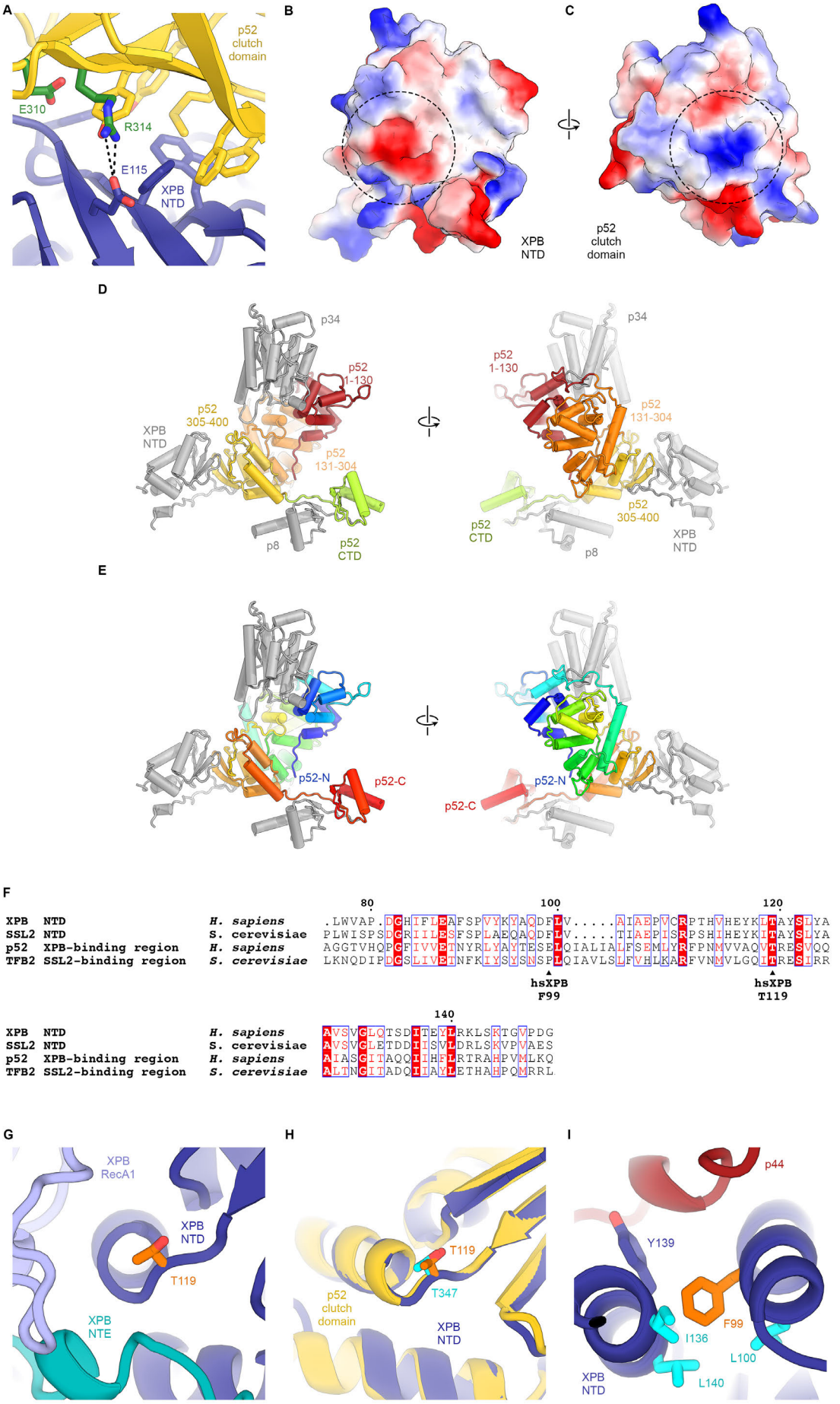


Figure 3 – figure supplement 1. Domain organization of XPB and sequence alignment of the N-terminal regions of XPB or XPB-like enzymes from eukaryotes, archaea, and bacteria. (**A**) Domain organization of human XPB, color-coded as in Figure 2C. (**B**) Sequence alignments. Sequences used (species and UNIPROT or GENBANK accession code indicated): Human (*Homo sapiens*, P19447), mouse (*Mus musculus*, P49135), fruit fly (*Drosophila melanogaster*, Q02870), Choanoflagellate (*Monosiga brevicollis*, A9V0A0), fungus

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(*Saccharomyces cerevisiae*, Q00578), ciliate (*Paramecium tetraurelia*, XP_001448651.1), plant (*Arabidopsis thaliana*, Q38861), archaea (*Ferroplasma*, EQB73227.1, *Thermoplasmatales*, EQB71900.1), bacteria (*Mycobacterium leprae*, Q9CBE0) (Balasingham et al., 2012; Poterszman et al., 1997). Some archaeal XPBs, including the *Archaeoglobus fulgidus* homolog that served for structure determination (Fan et al., 2006), contain only the DRD, lacking the NTD/NTE, and were not included. Residues are numbered according to the human sequence. Identical residues are printed white on red background, similar residues in red on white background. Secondary structure elements, residues mutated in human disease (F99S, T119S), and the approx. 70-residue insertion in the DRD are indicated. Figure generated using the ESPript web server (Robert & Gouet, 2014).

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



1397

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 3 – figure supplement 2. Structure and interactions of the XPB DRD and NTD. (**A**) Interface between the XPB NTD and the p52-clutch domains. Contacts include several large hydrophobic residues, as well as a salt bridge between p52 R314 and XPB E115 (shown as sticks). Mutations of residues equivalent to R314 and E310 in *D. melanogaster* p52 lead to disease-like phenotypes (Fregoso et al., 2007). (**B**, **C**) Electrostatic surface potential of the interacting surfaces of the XPB NTD (**B**) and the p52 clutch domain (**C**) show charge complementarity (strongest complementary charge peak indicated by a circle). (**D**) Front and back views of the domain organization of p52. Residues 1-130 (red) and 131-304 (orange) are not in contact with the XPB NTD. Residues 304-400 of p52 form the p52 clutch. Protein subunits or domains interacting with p52 are shown in grey. (**E**) Same as **D**, but p52 is colored using a gradient from blue (N-terminus) to red (C-terminus). (**F**) Sequence alignment of human and yeast p52 clutch and XPB NT domains highlights several highly conserved residues, including XPB T119, which is affected by a human disease mutation. Generated using the ESPript web server (Robert & Gouet, 2014). (**G**) Location of XPB T119, affected by the TTD mutation T119P, at a junction between a β-strand and an α-helix near the interface between XPB NTD, NTE, and RecA1 domains. (**H**) Superposition of the XPB NTD and the p52 clutch shows that a threonine at the position of XPB T119 also occurs in the p52 clutch. (**I**) The XP/CS mutation F99S affects a XPB NTD residue (orange) situated in a hydrophobic pocket formed by aliphatic residues (cyan) that are conserved across yeast and human XPB and p52 (**F**).

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

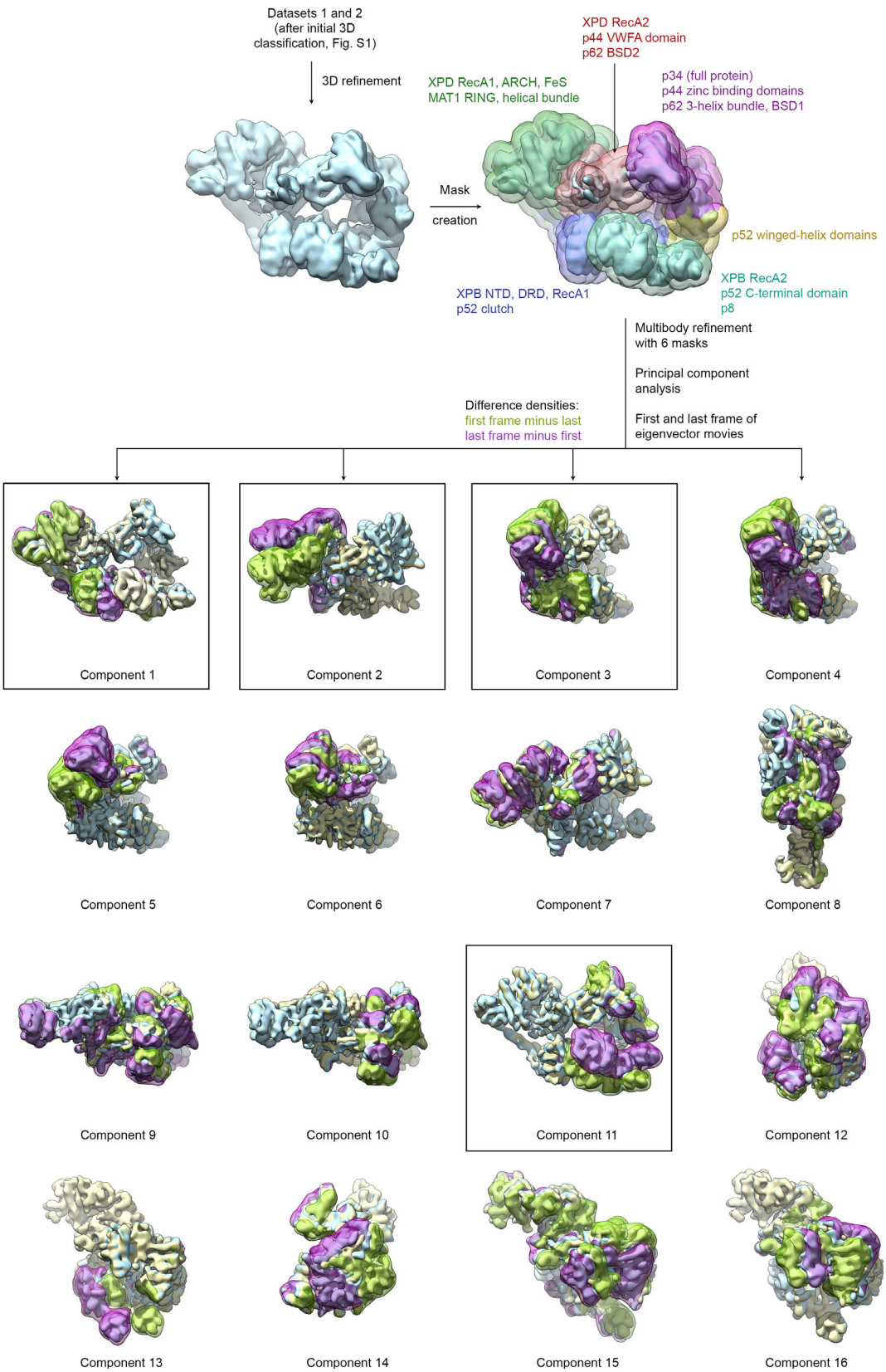


Figure 4 – figure supplement 1. Analysis of conformational variance of TFIIH. VPP datasets 1 and 2 were refined and subjected to multibody refinement using 6 masks followed by principal component analysis. The first and last

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

frames of the resulting eigenvector movies of the first 16 principal components are shown (in light blue and light yellow), along with difference densities obtained by subtracting the last volume from the first (green) and vice versa (purple). For the boxed components, the particles with the highest and lowest eigenvalues were extracted and refined (see Figure 4A).

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

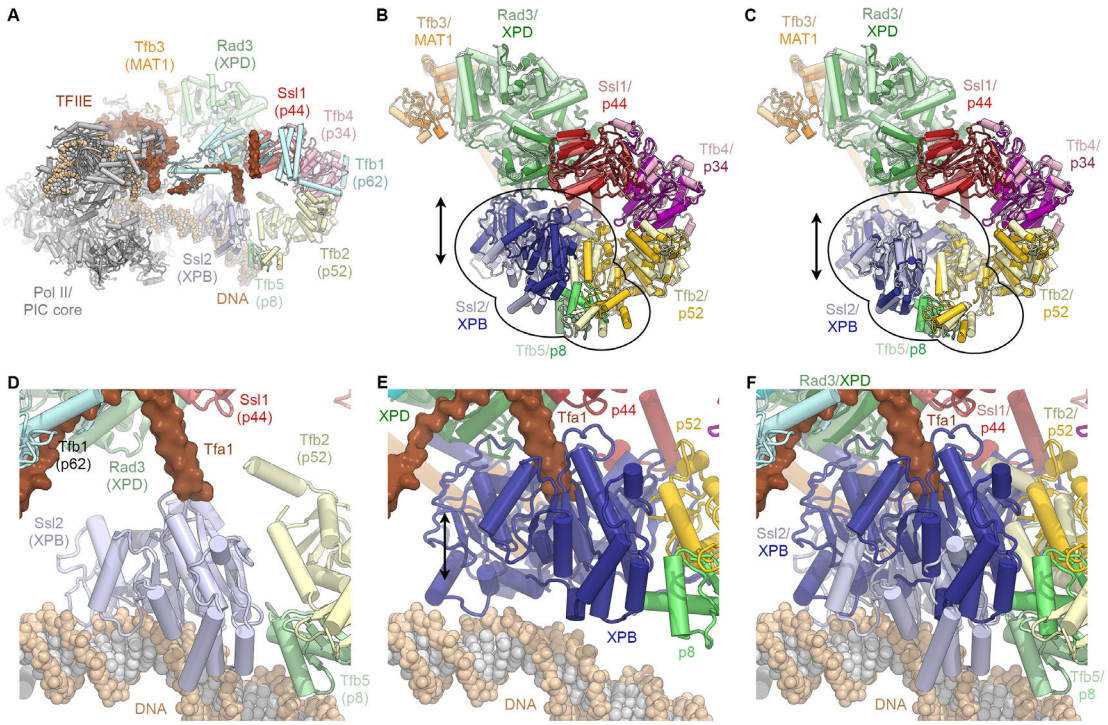


Figure 4 – figure supplement 2. Conformational dynamics of TFIIH and comparison with the PIC-bound TFIIH. (**A**) Rendering of the yeast TFIIH-containing Pol II-PIC (Schilbach et al., 2017). TFIIH subunits are colored and labeled with their yeast names, with human names in brackets. (**B**) Superposition of TFIIH from the yeast PIC (as shown in A, light colors) (Schilbach et al., 2017) and the complete structure of the human TFIIH core complex (this work; superposed on p44, bright colors) as single bodies. The top lobe of TFIIH (MAT1 helical bundle, XPB, p44, p34, p52 winged-helix domains; p62 not shown for clarity) aligns well, while the XPB-p52 (clutch/C-terminal domain)-p8 module is in a different position for the free and PIC-bound TFIIH (indicated). (**C**) Same as **B**, but with the human XPB-p52 (clutch/C-terminal domain)-p8 module superposed on the yeast structure as a rigid body. (**D**) View of DNA-bound Ssl2 (XPB) and neighboring subunits in the yeast Pol II-PIC (Schilbach et al., 2017). (**E**) View of human XPB, superposed on the yeast structure as in **B**. XPB could not bind DNA in this conformation, and the Tfa1 (TFIIE α-subunit) bridge helix, as observed in the yeast Pol II-PIC, clashes with XPB. (**F**) Panels **D** and **E** combined, visualizing the conformational change of TFIIH upon entry into the Pol II-PIC.

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

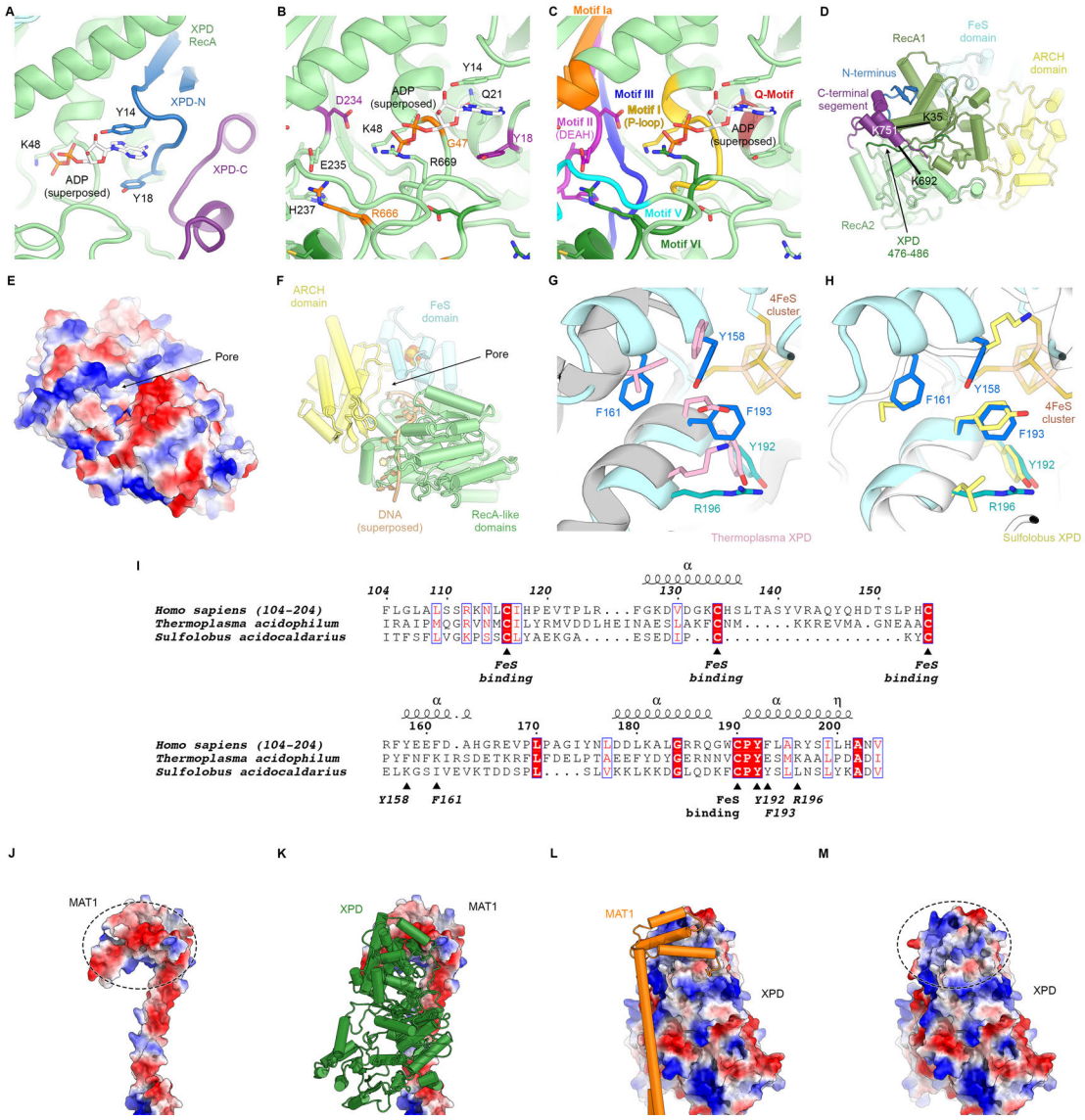


Figure 5 – figure supplement 1. Structure of XPD. (**A**) Close-up view of the XPD N- and C-terminal segments near the nucleotide binding pocket of the enzyme. Y12 and Y18 may stabilize nucleotides when bound (not present in our structure). (**B**) View of the nucleotide binding site of XPD with superposed ADP from nucleotide-bound DinG (PDB ID 6FWS (Cheng & Wigley, 2018)). Residues mutated in human disease are shown as sticks and colored (purple: XP, orange: XP/CS, dark green, TTD). These include Y18, which we propose could form a stacking interaction with the base of a bound nucleotide, as well as residues in conserved helicase motifs (see **C**). (**C**) Same view as **B**, but with conserved helicase motifs color-coded. Motifs Q, I, II, and IV are involved in nucleotide binding and hydrolysis. (**D**) View of the back side of XPD, with the N- and C-terminal segments indicated. The C-terminal segment interacts with the linker

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

between XPD RecA1 and RecA2 (residues 476-486). CX-MS crosslinks support the assignment of the C-terminal segment. The K751Q polymorphism has been associated with cancer therapy outcomes (Mlak et al., 2018; Peters et al., 2014), though functional differences between the variants has been debated (Clarkson & Wood, 2005). (**E**, **F**) Electrostatic surface (E) and cartoon (F) representations of the XPD structure, highlighting the pore between the XPD ARCH an FeS domains through which DNA (superposed from PDB ID 6FWR (Cheng & Wigley, 2018)) passes while being translocated by XPD. (**G**, **H**) Comparison of the structures of the human and archaeal FeS domains (**F**, *Thermoplasma*, PDB ID 4A15 (Kuper et al., 2012), **G**, *Sulfolobus*, PDB ID 3CRV (Fan et al., 2008)). There are differences at the secondary structure level, and functionally important residues are only partially conserved. (**I**) Sequence alignment of FeS domains from human, *Sulfolobus*, and *Thermoplasma* XPD. With the exception of the conserved cysteine residues that bind the FeS-cluster, sequence conservation in this domain is relatively low. Residues discussed in the text are indicated. (**J**-**M**) Charge complementary of regions of negative and positive electrostatic potential near the MAT1 3-helix bundle and the XPD ARCH domain interface (circled).

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

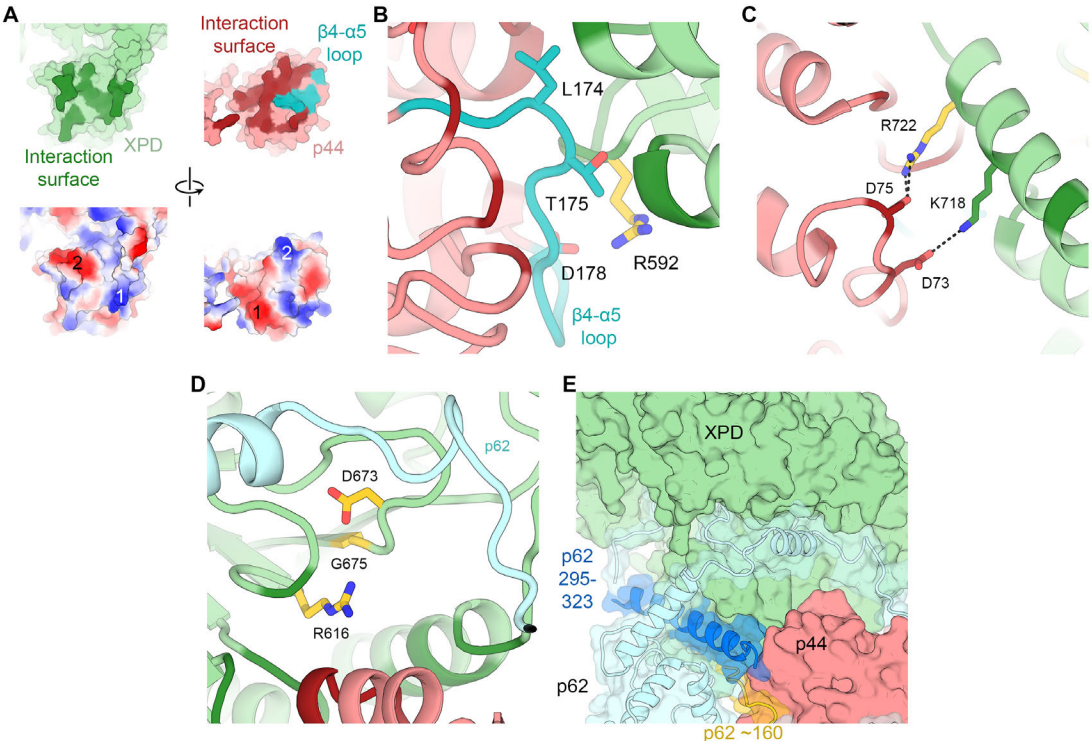


Figure 5 – figure supplement 2. XPD-p44 interaction. (**A**) Top: XPD-p44 (XPD green, p44 red) interaction surfaces are shown in darker shade (interacting residues in p44 β4-α5 loop are shown in teal). Bottom: Charge complementarity of the interface (two major complementary regions numbered 1 and 2). (**B**) Detailed view of the β4-α5 loop (teal) with mutations discussed in the text. XPD-p44 interacting regions (defined as residues within < 4 Å of neighboring protein) are colored in dark green (XPD) and dark red (p44). Residues discussed in the text are shown as sticks; those with mutation data (natural variants or experimental constructs) are colored yellow on XPD, teal on p44. (**C**) Detailed view of the interactions formed by XPD R722, which is mutated in TTD. (**D**) View of XPD residues R616, D673, and G675 in the context of p62 and p44. (**E**) XPD-p44-p62 interaction region. p62 segments interacting with both XPD and p44 are colored dark blue and yellow.

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

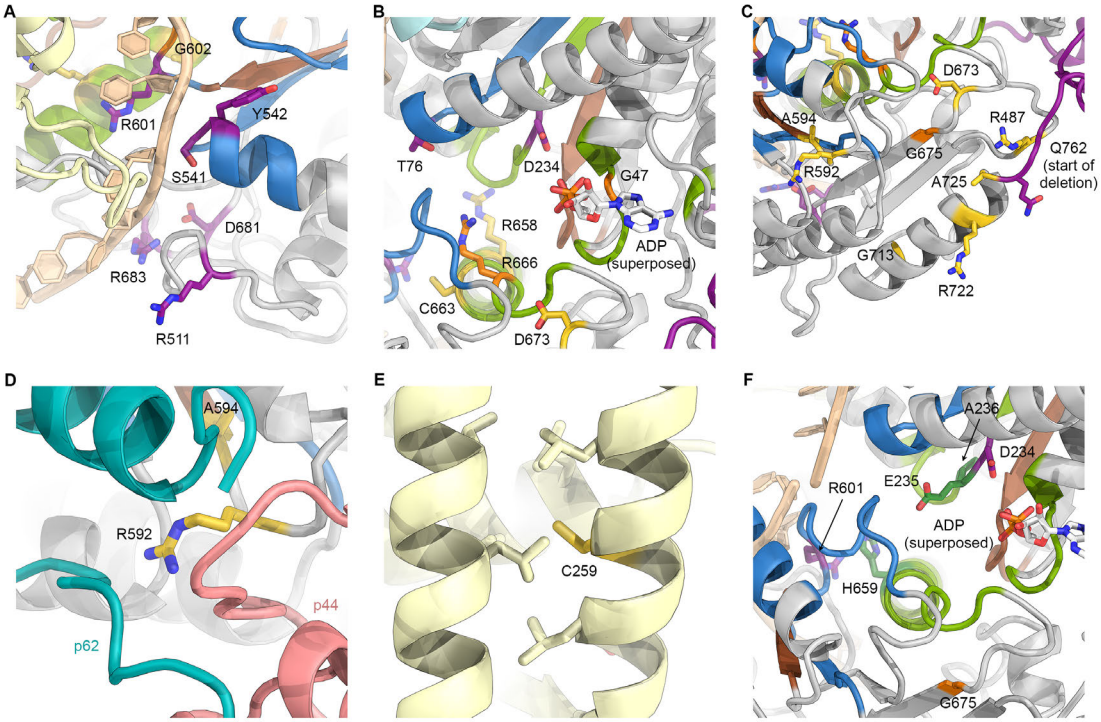


Figure 6 – figure supplement 1. Mapping of XPD disease mutations on the TFIIH structure. (**A**) Several residues affected by XP mutations localize near the DNA binding surfaces of XPD. This figure uses the same color scheme as Figure 6 (XP mutations purple; TTD yellow; CS-XP orange; helicase elements for DNA binding in blue, nucleotide binding and hydrolysis in green, coupling of nucleotide hydrolysis to DNA translocation in brown; DNA superposed from (Cheng & Wigley, 2018)). (**B**) Mapping of mutations near the catalytic site of XPD. ADP superposed from (Cheng & Wigley, 2018) (**C**) Mutations near the interface with p44. Note the predominance of TTD mutations in this peripheral region of XPD. Also see Figure 5E, Figure 5 – figure supplement 2, and text for details. (**D**) XPD R592, affected by a TTD mutation is in a key location where p44, p62, and XPD interact. (**E**) Location of C259, affected by the C259Y mutation, which likely causes disruption of ARCH domain integrity and leads to defects in CAK subcomplex recruitment. (**F**) Mapping of yeast Rad3 *rem* mutations (green) onto our XPD structure. E235 and A236 are part of the conserved DEAH sequence motif (helicase motif II), as is the immediately adjacent XP mutation D234. H659 localizes near DNA (according to the superposed model) and may be involved in ordering R601, which is in direct contact with the modeled DNA and affected by an XP mutation.

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Information

for

The complete structure of the human TFIIH core complex

Basil J. Greber, Daniel B. Toso, Jie Fang, Eva Nogales*

* Correspondence should be addressed to E.N. (ENogales@lbl.gov).

This PDF files includes:

Supplementary Tables 1-3
Supplementary Notes 1-2

Other Supplementary Information for this manuscript includes the following:

Supplementary Data Table S1

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Table 1. Data collection statistics. All datasets were acquired on Gatan K2 Summit direct electron detectors mounted in 300 kV-electron microscopes with three-condenser type electron optics. Abbreviations: TEM, transmission electron microscope; VPP, volta phase plate; Σ, sum.

Data set #	TEM	VPP	Pixel size (Å)	Total dose ($e^-/Å^2$)	Frames	K2 acquisition mode	Total movies	Movies retained	Particles after initial 3D sorting	Particles in final map	Data reported
1	Titan KRIOS	Yes	1.15	50	33	Super-resolution	3,270	2,047	154,298	27,165	This study
2	Titan KRIOS	Yes	1.15	50	33	Super-resolution	3,190	1,804	134,654	23,034	This study
3	Titan KRIOS	Yes	1.15	50	33	Super-resolution	4,804	3,269	191,830	18,365	This study
4	Titan KRIOS	Yes	1.15	50	33	Super-resolution	3,283	2,709	107,390	32,008	This study
5	Titan KRIOS	Yes	1.15	50	33	Super-resolution	4,958	2,857	85,582	22,717	This study
6	Titan KRIOS	Yes	1.15	50	50	Super-resolution	1,932	1,631	113,001	15,370	This study
7	Titan KRIOS	No	1.32	40	30	Super-resolution	2,719	476	12,788	-	This study
8	Titan	No	1.32	40	30	Counting	1,734	215	3,942	-	Greber et al. 2017
9	Titan	No	1.32	40	30	Counting	2,694	262	4,617	-	Greber et al. 2017
10	Titan	No	1.32	40	30	Counting	2,593	653	19,642	-	Greber et al. 2017
Σ							31,177	15,923	827,744	138,659	

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1561 **Supplementary Table 2. Components of the TFIIH structure.**

Protein	Chain ID	Size* (aa)	Modeled residues	Sequence accession code	Alternative names	Comments
XPB	A	782	34-204, 247-730	P19447	ERCC3	-
XPD	B	760	1-760	P18074	ERCC2	4Fe4S-cluster: SF4, residue 1000.
p62	C	548	107-173, 183-323, 345-547	P32780	GTF2H1	BSD1 domain docked and refined with reference restraints; residues 233-295 built as poly-alanines and deposited as UNK.
p52	D	462	7-458	Q92759	GTF2H4	
p44	E	395	15-267, 282-387	Q13888	GTF2H2	Zn^{2+}, residues 401, 402, 403
p34	F	308	8-72, 95-292	Q13889	GTF2H3	Zn^{2+}, residues 401
p8	G	71	2-67	Q6ZYL4	GTF2H5, TTDA	-
CDK7	-	346	-	P50613	MO15	Not visualized.
Cyclin H	-	323	-	P51946	CCNH	Not visualized.
MAT1	H	309	1-210	P51948	MNAT1	N-terminal RING domain docked and refined with tight reference restraints. Zn^{2+}, residues 400, 401

1562 * According to Uniprot (http://www.uniprot.org).

1563 † Unassigned sequences were initially modeled as poly-alanine and deposited as UNK.

1564

1565

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Table 3. Data collection, map and model refinement, model validation.

Data collection	**Datasets 1-6**
Microscope	Titan KRIOS
Stage type	Autoloader
Voltage (kV)	300
Detector	K2 Summit
Energy filter	GIF Quantum
Other	Volta phase plate
Pixel size (Å)	1.15
Defocus range (μm)	0.3-1.2
Electron dose (e^-/A^2)	50
Reconstruction	**EMD-XXXX**
Software	RELION 3
Particles	138,659
Box size (pixels)	256 x 256 x 256
Accuracy rotations (°)	1.4
Accuracy translations (pixels)	0.8
Map resolution (Å)	3.7
Map sharpening b-factor ($Å^2$)	-142
Coordinate refinement	
Software	PHENIX
Algorithm	REAL SPACE REFINE
Box size (Å)	294.4 x 294.4 x 294.4
Resolution cutoff (Å)	3.7
$FSC_{model\text{-}vs\text{-}map}$=0.5 (Å)	3.9
Model	**PDB-XXXX**
Number of residues	3203
Protein	3196
Ligand (Fe_4S_4, Zn^{2+})	7
B-factors overall	74.5
Protein	74.5
Ligand (Fe_4S_4)	63.4
Ligand (Zn^{2+})	133.8
R.m.s. deviations	
Bond lengths (Å)	0.009
Bond angles (°)	1.190
Validation	
Molprobity score (percentile)	2.0 (99th)
Molprobity clashscore (percentile)	6.7 (97th)
Rotamer outliers (%)	0.3
C_β deviations (%)	0
Ramachandran plot	
Favored (%)	87.8
Allowed (%)	11.9
Outliers (%)	0.3

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Discussion

Supplementary Note 1: The XPD-p44 interface. XPD serves a dual role within TFIIH: Its helicase activity is essential for TFIIH function in NER (Kuper et al., 2014; Winkler et al., 2000), and it serves as an architectural element in transcription, enabling the proper placement of the CAK subcomplex within the Pol II PIC (Abdulrahman et al., 2013; Drapkin et al., 1996; Dubaele et al., 2003; Greber et al., 2017; Guzder et al., 1994; He et al., 2016; Murakami et al., 2015; Robinson et al., 2016; Rossignol et al., 1997; Schilbach et al., 2017; Tirode et al., 1999; Tsai et al., 2017). The XPD-p44 interaction is critical both for the enzymatic and architectural roles of XPD (Coin et al., 1998; Dubaele et al., 2003; Kim et al., 2015; Kuper et al., 2014). Therefore, disruption of the XPD-p44 interface can simultaneously cause NER defects and impair transcription initiation, which leads to a TTD phenotype in human patients (Dubaele et al., 2003). Our structure provides a detailed view of this interaction interface, which encompasses the p44 loops β1-α1, β4-α5, β5-α6, and β6-α7 connecting β-strands and α-helices in the p44 VWA domain (Figure 5E, Figure 5 – figure supplement 2). Mutations L174W, T175R, and D178A in the p44 β4-α5 loop have been shown to abrogate the p44-XPD interaction and to reduce the helicase activity of XPD, implicating this region in XPD activation (Kim et al., 2015; Seroz et al., 2000). Our structure suggests that the insertion of large residues (L174W, T175R) into the tightly packed p44-XPD interface (Figure 5 – figure supplement 2B) is likely to lead to steric clashes, while the D178A mutation would break a likely salt bridge with XPD R592 (Figure 5C). Strikingly, the R592P mutation in XPD, the side chain of which packs against the p44 β4-α5 loop in wild type XPD, causes TTD in human patients (Cleaver et al., 1999), underscoring the importance of this interaction.

In XPD, RecA2 loop residues 531-533, 563-565, 589-592, 612-616, and the helical segment 718-725 form the p44 interaction interface (Figure 5E). Using recombinant human and *Chaetomium thermophilum* proteins, it has been shown that the XPD R722W mutation impairs the interaction between p44 and XPD, thereby interfering with both the p44-dependent stimulation of the XPD ATPase

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and helicase activities in NER, and architectural role of p44 in transcription initiation (Kuper et al., 2014). Our structure shows that the R722W mutation disrupts a salt bridge to p44 D75, and the bulky tryptophan side chain may additionally cause steric clashes with neighboring p44 residues, explaining the deleterious effect of the mutation (Figure 5E, Figure 5 – figure supplement 2C). It has been proposed that the R616P, D673G, and G675R disease mutations, which impair the p44-XPD interaction and the XPD helicase activity (Dubaele et al., 2003), exert their deleterious effect by impairing the interaction between XPD and p62 (Luo et al., 2015). However, our structure shows that G675 is buried in the enzymatic core (Figure 5 – figure supplement 2D), where accommodation of the large arginine side chain in G675R will disrupt the structure of the protein. XPD D673 is located in a β-strand in the vicinity of the XPD-p62 interface (Figure 5 – figure supplement 2D); however, no direct interactions of this residue with p62 are observed. XPD R616 is located at the p44-XPD interface, where its side chain is packed against neighboring residues and may engage in a salt bridge with p44 D204 (Figure 5 – figure supplement 2D). These interactions would be disrupted in the R616P mutant. Altogether, these observations suggest that these mutations act either via disruption of XPD structure or the XPD-p44 interface, not via p62. Nevertheless, a contribution of p62 to the stabilization of the p44-XPD interface cannot be excluded because several p62 elements are observed near the interface (Figure 5 – figure supplement 2E). As discussed in the main text, some of these interactions are dynamic (Figure 4 – figure supplement 2G-J) and unlikely to be strictly required for the stability of the p44-XPD interface.

Supplementary Note 2: Mapping of XPD mutations. In addition to our analysis of the XPD-p44 interface and the mutations located in its vicinity, we were also able to map numerous other human disease mutations onto our structure (Figure 6). Our results are in good agreement with the hypotheses derived from comparative modeling based on homologous bacterial DNA repair enzymes (Bienstock et al., 2003), archaeal XPD homologs (Fan et al., 2008; Liu et al., 2008; Wolski et al., 2008), and our previous lower-resolution TFIIH structure (Greber et al., 2017). Several XP mutations map in the regions that interact with bound DNA, as confirmed by superposing homologous DNA-bound helicase structures

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(Figure 6 – figure supplement 1A), or in the conserved motifs near the ATPase active site (Figure 6 – figure supplement 1B). These mutations impair XPD helicase activity (Dubaele et al., 2003; Fan et al., 2008; Liu et al., 2008), causing the NER defects that manifest in the XP disease symptoms. As explained above (Supplementary Note 1), TTD mutations tend to map to the periphery of XPD, including the XPD-p44 interface (Supplementary Note 1, Figure 5E, Figure 5 – figure supplement 2, Figure 6 – figure supplement 1C, D), the ARCH domain (Figure 6 – figure supplement 1E), or the FeS domain (Figures 5B, 6). The C259Y mutation in the ARCH domain likely impairs the proper folding of this domain due to steric hindrance introduced by the large tyrosine side chain, which in turn impairs association of the CAK subcomplex through disruption of MAT1 association, as discussed previously (Greber et al., 2017; Liu et al., 2008; Wolski et al., 2008). Consistent with its observed location in our structure (Figure 5B), the R112H mutation in the FeS domain leads to destabilization of the 4FeS cluster (Rudolf et al., 2006), which in turn leads to the disruption of folding of this entire domain (Fan et al., 2008) and lower TFIIH levels overall (Botta et al., 2002). Another TTD mutation that does not localize to the periphery of XPD is the particularly interesting temperature-sensitive mutation R658C, for which patients show exacerbated symptoms, including complete hair loss, when they develop a fever (Vermeulen et al., 2001). This residue is located at the interface between the XPD RecA1 and RecA2 domains, where it forms a salt bridge across the domain interface (Figure 6B). This interface would be partially destabilized by the mutation at normal temperature and more severely disrupted at elevated temperature, leading to stronger disease symptoms.

We also mapped onto our structure the location of the so-called *rem* mutations in Rad3, the yeast homolog of XPD. These mutations increase the affinity of Rad3 for DNA, which leads to retention of Rad3/XPD on DNA lesions, incomplete NER reactions, and subsequent DNA breakage upon arrival of a DNA replication fork (Herrera-Moyano et al., 2014; Moriel-Carretero & Aguilera, 2010). We mapped the *rem* mutations Rad3-E236G (XPD E235), Rad3-A237T (XPD A236), and Rad3-H661Y (XPD-H659Y), as well as the human XP-CS mutation XPD-G675R, which also exhibits elevated DNA affinity but reduced helicase activity in an

bioRxiv preprint doi: https://doi.org/10.1101/507723; this version posted January 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

archaeal model system (Fan et al., 2008) and DNA breakage in patient cells (Theron et al., 2005), onto our structure (Figure 6 – figure supplement 1F). XPD E235 and A236 are part of the conserved helicase motif II (signature sequence DEAH, Figure 5 – figure supplement 1C, Figure 6 – figure supplement 1F) (Fairman-Williams et al., 2010), and as such critical for nucleotide hydrolysis. Notably, the neighboring XPD D234, also part of motif II (Figure 5 – figure supplement 1C, Figure 6 – figure supplement 1C, F), is a human XP mutation (Cleaver et al., 1999). XPD H659 is located nearby, and contacts R601, which locates to the immediate vicinity of the DNA based on superposition with structures of DNA-bound helicases (Büttner et al., 2007; Cheng & Wigley, 2018) (Figure 6 – figure supplement 1F) and is affected by an XP mutation (Cleaver et al., 1999). In agreement with previous proposals, the *rem* mutations are therefore localized near regions responsible for ATP hydrolysis and DNA binding, consistent with the idea that they lead to stalling of NER complexes at DNA lesions, which may ultimately lead to incomplete NER, DNA breaks, failure to re-start transcription, and the XP-CS phenotype (Moriel-Carretero et al., 2015).

In summary, our structure pinpoints the locations of XPD mutations that provide insight into human disease mechanisms, and in most cases resolves the density for the side chains of the affected residues, thereby providing a structural framework for the detailed analysis of the interactions of the affected residues and the effects of these mutations.

Analysis of CX-MS data for human TFIIH

Chemical crosslinking-mass spectrometry data taken from Luo et al., Mol. Cell 59 (5): 794-806 (2015) and mapped onto the cryo-EM structure of human TFIIH.

Residues not built in in the cryo-EM structure of human TFIIH are printed red.
Crosslinking distances refer to Cα-Cα distance of crosslinked lysine residues.
The expected maxium crosslink distance for BS3-induced crosslinks is 30 Å.

Color coding for crosslinking distances:	Agreement (< 30 Å)	Borderline (30-35 Å)	Incompatible (> 35 Å)

Intramolecular crosslinks in human TFIIH subunits

Protein	Residue 1	Residue 2	Distance (Å)	Comments
p34	100	56	18.0	
p34	149	56	14.0	
p34	97	56	23.2	
p34	74	56	-	
p34	152	56	14.6	
p34	125	56	35.2	
p34	125	74	-	
MAT1	127	137	16.8	
MAT1	167	180	19.6	
MAT1	176	202	11.9	
MAT1	55	70	16.0	
MAT1	137	55	27.3	
MAT1	70	76	10.0	
MAT1	70	77	10.7	
XPD	268	274	9.7	
XPD	128	603	20.5	
XPD	692	751	28.6	
XPD	223	228	9.3	
XPD	223	69	13.0	
XPD	223	445	14.0	
XPD	223	35	16.8	
XPD	133	181	18.4	
XPD	35	751	26.8	
XPD	603	82	17.5	
XPD	228	35	10.4	
XPD	682	689	10.1	
XPD	113	603	18.6	
XPD	113	128	9.3	
XPD	113	181	11.3	
XPD	507	634	19.5	
XPD	507	682	9.8	
CDK7	32	342	-	CKD7 is not modeled in our structure.
CDK7	44	52	-	
CDK7	328	42	-	
CDK7	14	42	-	
CDK7	14	32	-	
CDK7	14	44	-	
CDK7	330	343	-	
XPB	167	222	-	Crosslink of the DRD loop insertion
XPB	167	326	16.3	
XPB	569	715	12.4	
XPB	171	311	29.2	
XPB	171	217	-	Crosslinks of the DRD loop insertion
XPB	171	222	-	that is not visible in the cryo-EM map
XPB	44	778	-	
XPB	44	68	28.1	
XPB	322	326	6.2	
XPB	157	44	13.1	
XPB	549	577	12.0	
XPB	282	463	18.0	
XPB	282	476	21.5	
XPB	577	601	11.3	
XPB	142	222	-	Crosslinks of the DRD loop insertion
XPB	142	241	-	that is not visible in the cryo-EM map
XPB	142	68	12.1	
XPB	311	379	10.0	
XPB	311	326	23.2	
XPB	217	326	-	Crosslinks of the DRD loop insertion
XPB	222	326	-	that is not visible in the cryo-EM map
XPB	222	257	-	
p62	129	504	15.3	
p62	18	38	-	Crosslinks of the N-terminal PH-
p62	104	106	-	domain of p62 and other N-terminal
p62	19	38	-	segments, which are not modeled
p62	161	38	-	in our structure.
p62	102	106	-	
p62	102	109	-	
Cyclin H	253	260	-	Cyclin H is not modeled in our
Cyclin H	274	62	-	structure.
Cyclin H	293	295	-	
Cyclin H	189	72	-	
Cyclin H	62	72	-	
Cyclin H	292	295	-	
p44	116	126	13.3	
p44	116	306	20.1	
p44	116	157	15.6	
p44	23	40	18.2	
p44	126	23	55.5	
p44	126	306	16.9	
p44	126	133	10.4	
p44	126	268	28.6	
p44	126	297	20.4	
p52	76	81	10.1	
p52	452	459	-	
p52	452	456	6.7	
p52	378	53	31.8	
p52	127	53	24.6	
p8	2	31	24.5	
p8	6	31	17.0	

Intermolecular crosslinks in human TFIIH subunits

Protein 1	Residue 1	Protein 2	Residue 2	Distance (Å)	Comments
p34	149	p52	378	15.3	
p34	152	p52	378	10.2	
p34	97	p52	127	8.7	
XPB	379	XPD	682	29.7	
XPB	222	XPD	682	-	
XPB	326	XPD	682	18.4	
XPB	171	XPD	682	19.6	
XPB	171	XPD	507	19.1	
p62	378	p34	125	25.7	
p62	38	p34	125	-	
p62	38	XPB	549	-	
p62	353	p44	126	50.0	p62-p44 crosslinks were discarded
p62	38	p44	126	-	in modeling due to high incidence
p62	252	p44	31	73.1	of distance violations and excellent
p62	161	p44	116	22.3	agreement of p62 model with
p62	161	p44	126	34.0	p62-XPD crosslinks.
p62	350	XPD	82	10.2	
p62	350	XPD	113	17.0	
p62	365	XPD	96	16.9	
p62	353	XPD	603	19.6	
p62	353	XPD	113	20.8	
p62	353	XPD	82	9.4	
p62	275	XPD	603	13.6	
p62	293	XPD	603	23.0	
p62	293	XPD	113	32.1	
p62	296	XPD	603	25.7	
Cyclin H	239	MAT1	55	-	Cyclin H is not modeled.
MAT1	121	XPB	44	116.7	
MAT1	180	XPB	171	17.7	
MAT1	202	XPB	142	37.5	
MAT1	176	XPB	171	19.3	
MAT1	167	XPB	217	-	Crosslinks of the DRD loop insertion
MAT1	167	XPB	222	-	that is not visible in the cryo-EM map
MAT1	167	XPB	257	19.1	
MAT1	167	XPD	689	26.9	
MAT1	167	XPD	634	19.7	
p44	40	p34	149	21.0	
p44	126	p34	149	34.1	
p44	42	p34	149	22.2	
p44	229	p34	149	23.0	
p44	229	p34	152	26.7	
p44	23	XPB	142	7.7	
p44	31	XPB	142	8.7	
p44	23	p52	378	32.7	
p44	40	p52	378	23.2	
p44	268	p52	77	30.0	
p44	77	p52	378	20.8	
p52	438	XPB	526	33.9	
p52	438	XPB	706	20.4	
p52	378	XPB	526	51.6	
p52	378	XPB	241	-	
p52	378	XPB	257	63.0	
p52	140	XPD	29	92.4	Likely false positives given the reliable
p52	277	XPD	101	116.0	assignment of p52 and XPD.
p52	53	p8	31	59.3	
p52	452	p8	2	9.8	
p52	452	p8	6	12.5	
p52	77	p8	71		
p52	456	p8	2	11.8	
p52	456	p8	6	18.5	
CDK7	52	Cyclin H	293	-	CDK7 and Cyclin H are not modeled
CDK7	293	XPD	634	-	CDK7 is not modeled.
p8	71	XPB	767	-	Crosslinks of p8 with the XPB RecA2
p8	71	XPB	526	-	domain are compatible with the
p8	71	XPB	59	-	architecture of TFIIH overall.
p8	71	XPB	476	-	
p8	6	XPB	167	88.0	