

The cryo-EM structure of human TFIID

Basil J. Greber^{1,2}, Thi Hoang Duong Nguyen^{1,2,3}, Jie Fang⁴, Pavel V. Afonine², Paul D. Adams^{2,5}, Eva Nogales^{1,2,4,6,*}

¹ California Institute for Quantitative Biology (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 Miller Institute for Basic Research in Science, University of California, Berkeley, California 94720, USA

⁴ Howard Hughes Medical Institute, University of California, Berkeley, California 94720, USA

⁵ Department of Bioengineering, University of California, Berkeley, California 94720, USA.

⁶ Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA

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

Human transcription factor IIH (TFIIH) is part of the general transcriptional machinery required by RNA polymerase II (Pol II) for the initiation of eukaryotic gene transcription ¹. Composed of 10 subunits that add up to a molecular weight of about 500 kDa, TFIIH is also essential for nucleotide excision repair (NER) ¹. The seven-subunit TFIIH core complex formed by XPB, XPD, p62, p52, p44, p34, and p8 is competent for DNA repair ², while the CDK activating kinase (CAK) subcomplex, which includes the kinase activity of CDK7 as well as the Cyclin H and MAT1 subunits, is additionally required for transcription initiation ^{1,2}. Mutations in the TFIIH subunits XPB, XPD, and p8 lead to severe premature aging and cancer propensity in the genetic diseases xeroderma pigmentosum (XP), Cockayne's syndrome (CS), and trichothiodystrophy (TTD), highlighting the importance of TFIIH for cellular physiology ³. Here, we present the cryo-electron microscopy structure of human TFIIH at 4.4 Å resolution. The structure reveals the molecular architecture of the TFIIH core complex, the detailed structures of its constituent XPB and XPD ATPases, and how the core and kinase subcomplexes of TFIIH are connected. Additionally, our structure provides insight into the conformational dynamics and the regulation of TFIIH activity.

We collected cryo-EM data of human TFIIH, immuno-purified from HeLa cells, and obtained a reconstruction at an overall resolution of 4.4 Å (see Methods and Extended Data Figs. 1, 2). The reconstruction (Fig. 1a) shows a horseshoe-shaped assembly corresponding to the TFIIH core complex, in agreement with previous lower-resolution reconstructions of human TFIIH in isolation ⁴ and in the context of the Pol II-PIC ^{5,6}. Secondary structure elements are resolved throughout (Extended Data Fig. 3a-c), and large amino acid side chains are visible in the best-resolved areas of the map (better than 4 Å; Extended Data Fig. 2b, 3d-f). Because the resolution of our cryo-EM map was not sufficient for chain tracing and assignment of the sequence register in several areas, we combined docking and rebuilding of homology models with the placement of secondary structure elements for previously structurally uncharacterized components in order to generate an atomic

model of the TFIIH core-MAT1 complex (Fig. 1b, Extended Data Table 1). This atomic model was refined against the map and fully validated (Extended Data Fig. 3g-i). All protein subunits of the TFIIH core complex except p62, as well as the CAK component MAT1, could be unambiguously assigned in our cryo-EM map, while the remainder of the CAK subcomplex is highly flexible and its CDK7 and Cyclin H subunits are not resolved (Fig. 1a, b). The architecture of TFIIH is dominated by its ATPase/helicase subunits XPD and XPB, which are situated next to each other at the open end of the horseshoe-shaped structure (green and blue, respectively, in Fig. 1a, b). The arc-like connection between them includes the von Willebrand Factor A (vWFA)-like folds of p44 and p34 and the helical domain of p52 (red, magenta, and yellow, respectively, in Fig 1a, b; domain architectures are shown in Extended Data Fig. 4). Even though p62 could not be unambiguously assigned in our cryo-EM map, data from chemical crosslinking-mass spectrometry (CX-MS) experiments ⁷ suggest that a number of secondary structure elements built into unassigned density may correspond to parts of p62 (Fig. 1c, Extended Data Fig. 5, Supplementary Discussion). These α -helical regions and the linkers that seem to connect them localize to distant parts of TFIIH, suggesting that p62 may act as a molecular glue to stabilize the overall assembly of TFIIH.

The structures of XPD and XPB (Fig. 2) show that both ATPases are in the apo-state, with unoccupied nucleotide binding pockets. Human XPD shares its overall architecture with homologous archaeal helicases ⁸⁻¹⁰, with two RecA-like domains (RecA1 and RecA2) harboring the structurally conserved helicase motifs, and 4FeS and ARCH domain insertions within RecA1 (Fig. 2a, Extended Data Fig. 4). The 4Fe4S cluster is clearly visualized in the cryo-EM map as a strong density peak within the 4FeS domain (Extended Data Fig. 6a-c) ^{8,10}. The 4FeS domain is adjacent to the pore, through which single stranded DNA has been proposed to be threaded before reaching the two RecA-like domains of XPD (Extended Data Fig. 6c-e) ^{11,12}. The region around the entrance to the pore in human XPD is narrower than in the archaeal enzymes due to a nearby insertion segment within the human ARCH

domain (Extended Data Fig. 6c, e), although the pore still appears to be accessible along the path proposed for archaeal XPD (Extended Data Fig. 6e)^{8,10}.

The helicase activity of XPD is known to be up-regulated by interactions with p44, which are also required for assembly of the helicase into the TFIIH complex¹³. Our structure shows that these interactions with p44 localize near the conserved helicase motifs IV and V in XPD RecA2 (Extended Data Fig. 6f), but on the outside of the RecA-like domain, facing away from the DNA binding site (Extended Data Fig. 6d, g). Mutations of XPD residues near this interaction surface (Fig. 2b) cause XP or TTD^{1,3,10}, highlighting the importance of the XPD-p44 interaction for assembly and function of TFIIH. Unlike proteins that stimulate the activity of DNA and RNA helicases by contacting their N-terminal RecA-like domain and promoting substrate release, such as the translation initiation factor eIF4G and the protein Gle1, which regulate the DEAD-box RNA helicases eIF4A and Dbp5, respectively (Extended Data Fig. 6g, h)^{14,15}, p44 is more likely to regulate XPD through subtle effects on the conformation of the nearby helicase motifs IV and V, which are involved in DNA binding and coupling of ATPase and helicase activity¹⁶. This hypothesis agrees with the observation that p44 enhances the helicase activity of XPD, but does not alter its ATPase activity¹³.

Mapping of disease-causing mutations³ onto our structure of human XPD shows that most of the mutations that cause XP or XP/CS cluster near the DNA- or ATP-binding sites of the RecA-like domains of XPD (Fig. 2c), in agreement with their strong effect on helicase activity in both human and archaeal XPD^{9,10,17}. Their position explains the defects they cause in NER¹⁷ and confirms previous predictions based on the structures of archaeal XPD homologs^{9,10}. While these XP or XP/CS mutations are mostly confined to the helicase core of the enzyme, TTD-causing mutations also affect peripheral regions important for contacts with the other components of TFIIH (Fig. 2b, c). Some of these TTD mutations are located at the interaction site with p44 (see above) and in the ARCH domain, which is the site of association with MAT1 (see below). TTD-causing mutations have been associated

with defects in both NER and basal transcription *in vitro* ¹⁷. Because transcription does not depend on the helicase activity of XPD ¹⁸ and the TTD mutations include sites away from the critical elements in the helicase domain, some of these mutations may cause transcription defects by impairing the ability of XPD to correctly assemble into TFIIH, where it may be required to support the association of TFIIH with the Pol II-PIC and to correctly position the CAK subcomplex near Pol II during transcription initiation (see below).

The structure of human XPB (Fig. 2d) encompasses the two RecA-like domains, a DNA-damage recognition domain (DRD)-like domain, and a set of secondary structure elements corresponding to the N-terminal extension domain (NTE). The NTE is important for anchoring and activity of XPB within TFIIH and affected by disease mutations ^{3,7,19}, but not present in archaeal XPB. Although a precise delineation of p52 and the XPB-NTE at their contact site is not possible at the present resolution, our structure suggests this functionally important interaction may involve regions of predicted β -strands within both proteins (Fig. 2e; for further discussion of the DRD and NTE see Extended Data Figs. 5, 7 and Supplementary Discussion). The overall structure of the two RecA-like domains is similar to the previously reported archaeal XPB structure, including the signature RED and thumb motifs, even though the latter is substantially shortened relative to its archaeal counterpart (Fig. 2d, f) ²⁰. XPB also shares structural homology with the archaeal ssoRad54 SWI2/SNF2-type ATPase ²¹ and the eukaryotic Rad54 SWI2/SNF2 chromatin remodeler ²², including the presence of a domain insertion at the site of the XPB thumb domain (Extended Data Fig. 6i) and conserved motifs in the RecA-like domains (Extended Data Fig. 6j) ¹⁶. The relative orientation of the two RecA-like domains observed in XPB within TFIIH is similar to the arrangement in the structure of eukaryotic Rad54 ²², which lacks bound substrate, but highly dissimilar to the conformations of both the archaeal ssoRad54 SWI2/SNF2 ATPase in complex with DNA ²¹ and archaeal XPB ²⁰ as reported previously ⁵. Comparison with human TFIIH within a promoter bound Pol II-PIC ⁵ shows that DNA binding does not induce major inter-domain rearrangements in XPB within TFIIH. In fact, the conformations

observed in DNA-bound ssoRad54 (ref. ²¹) or in free archaeal XPB would cause extensive steric hindrance or rearrangement of large interaction interfaces (Extended Data Fig. 6k). It seems that the structure of XPB within TFIIH is more restrained than for homologous enzymes that are not part of this large assembly and that the TFIIH scaffold may act to prearrange the otherwise flexible RecA-like domains of XPB in a conformation that is suitable for substrate binding and catalytic activity, in agreement with previous biochemical data ²³.

Similarly to the p44-XPB interaction, XPB RecA2 interacts with p52 and p8, interactions that are required for XPB assembly and function within TFIIH ^{7,19}. The L21P mutation in p8 that causes the disease TTD in humans ²⁴ is located in an α -helix that is sandwiched between the fold of p8 and XPB RecA2 (Fig. 2g). The presence of a proline at this site likely distorts or unfolds the affected helix, thereby destabilizing the interactions between XPB, p8, and p52 in this region or reducing p8 stability ²⁵.

In our cryo-EM map, an elongated density connects the DRD-like domain of XPB with a region of XPD RecA2 that is inferred to be part of the DNA substrate-binding site based on the structures of archaeal XPD-DNA complexes ^{11,12} (Fig. 3a). While this density could be due to residual DNA not removed during TFIIH purification from native source, an alternative possibility is that it corresponds to an unassigned protein region that could potentially act to regulate the interaction of XPD with substrate.

We assigned the CAK subunit MAT1 to a very long α -helix and a helical bundle that form interactions with both XPB and the XPD ARCH domain (Fig. 3a, b). This density localizes near the region of attachment of the CAK subcomplex to the TFIIH core, as visualized in 2D class averages and 3D reconstructions of negatively stained TFIIH (Extended Data Fig. 8a-c and refs. ^{6,26}), and reconstructions of TFIIH in the context of the assembled Pol II-PIC (see below and refs. ^{5,27}). Previous data additionally show that MAT1 interacts with the XPD ARCH domain, and that the C259Y mutation in the

ARCH domain impairs binding²⁸. Combined with secondary structure prediction showing a very long α -helix in MAT1 (Extended Data Fig. 4) and CX-MS and site-specific crosslinking data showing proximity of MAT1 to XPB and XPD (Fig. 3c, d)^{7,29}, these observations strongly support our assignment of MAT1. At the end of the MAT1 helical bundle, there is weak density that could accommodate the N-terminal RING domain of MAT1 (blue in Fig. 3d, Extended Data Fig. 8d, e), however unambiguous assignment is currently not possible.

MAT1 connects all three ATP-dependent moieties of TFIIH – the CAK subcomplex and the XPB and XPD ATPases – and may be involved in the regulation and coordination of these functional centers of TFIIH. Interestingly, the presence of the CAK subcomplex, which includes MAT1, is known to inhibit the activity of the XPD helicase^{28,30}. Accordingly, the CAK subcomplex is present during transcription initiation, when the XPD helicase is inactive, and removed during NER, when XPD helicase activity is required^{2,18,19,31}. The interactions between MAT1 and XPD observed in our cryo-EM map might be involved in mediating this inhibition, possibly by limiting the conformational freedom of the ARCH domain, which has been implicated in DNA substrate loading by XPD¹¹. In the context of the human Pol II-PIC⁵, MAT1 may also be involved in forming contacts between TFIIH and the PIC core (Fig. 3e). Furthermore, MAT1 is in close proximity to a density element located between TFIIH and the Pol II-PIC core that probably corresponds to the remainder of the CAK subcomplex (Fig. 3f)^{5,6}, according to our negative stain reconstructions of apo-TFIIH (Extended Data Fig. 8f, g and ref. 6) and in agreement with recent studies in yeast²⁷.

The XPD C259Y mutation, which is found in TTD patients³ and impairs both NER and transcription activity *in vitro*²⁸, likely destabilizes the structure of the entire ARCH domain^{8,9}. Our structure now indicates that perturbations of the structure of the ARCH domain may prevent functional incorporation of MAT1 into TFIIH, which will impair the proper placement of the remaining components of the CAK subcomplex within the Pol II-PIC, thus explaining the defects in Pol II-CTD

phosphorylation and transcription observed in the presence of the C259Y mutant TFIIH *in vitro* ²⁸.

Conformational changes in molecular machines are often associated with functional transitions. In free TFIIH, an interaction between XPB and XPD, possibly supported by additional XPB-p44 contacts, brings together the two ends of the horseshoe-shaped TFIIH core complex (Fig. 4a, Extended Data Fig. 9a). In the Pol II-PIC ⁵, TFIIH undergoes a conformational change during which the XPD-XPB interaction breaks as the distance between them is increased relative to free TFIIH, (Fig. 4b, c, Extended Data Fig. 9b-f). Overall, the RecA-like domains of XPB move away from XPD when engaging DNA. Because they are bound to XPB RecA2, p8 and the C-terminal domain of p52 are also relocated by this motion (Fig. 4c). These rearrangements may be required for the activation of the DNA translocase function of XPB within the Pol II-PIC that has been proposed to aid the melting of the promoter DNA and the formation of the transcription bubble ^{5,6,32}. Additionally, the MAT1 contact site on the XPD ARCH domain is clearly resolved in the reconstruction of the Pol II-PIC ⁵ (Fig. 3e, f), while the density for the connection to XPB is not clear. This might indicate that the MAT1-XPB contact is released in the Pol II-PIC to facilitate conformational changes that enable the remainder of the CAK subcomplex to access and phosphorylate its targets in the Pol II-PIC. Complete release of the CAK during NER ^{2,31} may not only de-repress the helicase activity of XPD ^{28,30}, but might also allow for more conformational flexibility of TFIIH to possibly facilitate the recruitment of downstream NER factors.

References

- 1 Compe, E. & Egly, J.-M. Nucleotide Excision Repair and Transcriptional Regulation: TFIIH and Beyond. *Annu. Rev. Biochem.* **85**, 265-290 (2016).
- 2 Svejstrup, J. Q. *et al.* Different forms of TFIIH for transcription and DNA repair: holo-TFIIH and a nucleotide excision repairosome. *Cell* **80**, 21-28 (1995).

- 3 Cleaver, J. E., Thompson, L. H., Richardson, A. S. & States, J. C. A summary of
mutations in the UV-sensitive disorders: xeroderma pigmentosum, Cockayne
4 syndrome, and trichothiodystrophy. *Hum Mutat* **14**, 9-22 (1999).
- 5 Schultz, P. *et al.* Molecular structure of human TFIIH. *Cell* **102**, 599-607
(2000).
- 6 He, Y. *et al.* Near-atomic resolution visualization of human transcription
promoter opening. *Nature* **533**, 359-365 (2016).
- 7 He, Y., Fang, J., Taatjes, D. J. & Nogales, E. Structural visualization of key steps
in human transcription initiation. *Nature* **495**, 481-486 (2013).
- 8 Luo, J. *et al.* Architecture of the Human and Yeast General Transcription and
DNA Repair Factor TFIIH. *Mol. Cell* **59**, 794-806 (2015).
- 9 Wolski, S. C. *et al.* Crystal structure of the FeS cluster-containing nucleotide
excision repair helicase XPD. *Plos Biol.* **6**, e149 (2008).
- 10 Liu, H. *et al.* Structure of the DNA Repair Helicase XPD. *Cell* **133**, 801-812
(2008).
- 11 Fan, L. *et al.* XPD helicase structures and activities: insights into the cancer
and aging phenotypes from XPD mutations. *Cell* **133**, 789-800 (2008).
- 12 Constantinescu-Aruxandei, D., Petrovic-Stojanovska, B., Penedo, J. C., White,
M. F. & Naismith, J. H. Mechanism of DNA loading by the DNA repair helicase
XPD. *Nucleic Acids Res.* **44**, 2806-2815 (2016).
- 13 Kuper, J., Wolski, S. C., Michels, G. & Kisker, C. Functional and structural
studies of the nucleotide excision repair helicase XPD suggest a polarity for
DNA translocation. *EMBO J* **31**, 494-502 (2012).
- 14 Coin, F. *et al.* 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 (1998).
- 15 Schütz, P. *et al.* Crystal structure of the yeast eIF4A-eIF4G complex: an RNA-
helicase controlled by protein-protein interactions. *Proc. Natl. Acad. Sci. U. S.*
A. **105**, 9564-9569 (2008).
- 16 Montpetit, B. *et al.* A conserved mechanism of DEAD-box ATPase activation
by nucleoporins and InsP6 in mRNA export. *Nature* **472**, 238-242 (2011).
- 17 Fairman-Williams, M. E., Guenther, U.-P. & Jankowsky, E. SF1 and SF2
helicases: family matters. *Curr. Opin. Struct. Biol.* **20**, 313-324 (2010).
- 18 Dubaele, S. *et al.* Basal transcription defect discriminates between xeroderma
pigmentosum and trichothiodystrophy in XPD patients. *Mol. Cell* **11**, 1635-
1646 (2003).
- 19 Kuper, J. *et al.* In TFIIH, XPD helicase is exclusively devoted to DNA repair.
Plos Biol. **12**, e1001954 (2014).
- 20 Coin, F., Oksenychn, V. & Egly, J.-M. 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 (2007).
- 21 Fan, L. *et al.* Conserved XPB core structure and motifs for DNA unwinding:
implications for pathway selection of transcription or excision repair. *Mol.*
Cell **22**, 27-37 (2006).

- 21 Dürr, H., Körner, C., Müller, M., Hickmann, V. & Hopfner, K.-P. X-ray structures of the *Sulfolobus solfataricus* SWI2/SNF2 ATPase core and its complex with DNA. *Cell* **121**, 363-373 (2005).
- 22 Thomä, N. H. *et al.* Structure of the SWI2/SNF2 chromatin-remodeling domain of eukaryotic Rad54. *Nat. Struct. Mol. Biol.* **12**, 350-356 (2005).
- 23 Grünberg, S., Warfield, L. & Hahn, S. Architecture of the RNA polymerase II preinitiation complex and mechanism of ATP-dependent promoter opening. *Nat. Struct. Mol. Biol.* **19**, 788-796 (2012).
- 24 Giglia-Mari, G. *et al.* A new, tenth subunit of TFIIH is responsible for the DNA repair syndrome trichothiodystrophy group A. *Nat Genet* **36**, 714-719 (2004).
- 25 Kainov, D. E., Vitorino, M., Cavarelli, J., Poterszman, A. & Egly, J.-M. Structural basis for group A trichothiodystrophy. *Nat. Struct. Mol. Biol.* **15**, 980-984 (2008).
- 26 Gibbons, B. J. *et al.* Subunit architecture of general transcription factor TFIIH. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 1949-1954 (2012).
- 27 Tsai, K.-L. *et al.* Mediator structure and rearrangements required for holoenzyme formation. *Nature* **544**, 196-201 (2017).
- 28 Abdulrahman, W. *et al.* ARCH domain of XPD, an anchoring platform for CAK that conditions TFIIH DNA repair and transcription activities. *Proc. Natl. Acad. Sci. U. S. A.* **110**, E633-642 (2013).
- 29 Warfield, L., Luo, J., Ranish, J. & Hahn, S. Function of conserved topological regions within the *S. cerevisiae* basal transcription factor TFIIH. *Mol. Cell Biol.* **36**, 2464-2475 (2016).
- 30 Sandrock, B. & Egly, J. M. 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 (2001).
- 31 Coin, F. *et al.* Nucleotide excision repair driven by the dissociation of CAK from TFIIH. *Mol. Cell* **31**, 9-20 (2008).
- 32 Fishburn, J., Tomko, E., Galburt, E. & Hahn, S. Double-stranded DNA translocase activity of transcription factor TFIIH and the mechanism of RNA polymerase II open complex formation. *Proceedings of the National Academy of Sciences* **112**, 3961-3966 (2015).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgments We thank S. Zheng and D. King for providing peptides and XPB mAb, and P. Grob, T. Houweling, and A. Chintangal for cryo-EM and computing support, respectively. We thank A. B. Patel and R. K. Louder for discussions and help with data collection, and J. H. D. Cate for providing PYMOL scripts. We acknowledge the use of the LAWRENCIUM computing cluster at Lawrence Berkeley National

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

Author contributions E.N. directed the study. J.F. performed HeLa cell culture and prepared TFIIH. B.J.G. performed cryo-EM specimen preparation, data collection, data processing, and initial model building. B.J.G. and T.H.D.N. built the final model. B.J.G. and T.H.D.N. performed coordinate refinement, supported by P.V.A. in the laboratory of P.D.A. B.J.G. wrote the initial draft of the manuscript and all authors contributed to the final version.

Author information Reprints and permissions information is available at www.nature.com/reprints. The authors do not declare any competing financial interests. Correspondence and requests for materials should be addressed to E.N. (enogales@lbl.gov).

Data availability The cryo-EM map of the TFIIH core complex has been deposited in the Electron Microscopy Databank (EMDB) with accession code EMD-3802 (final refined map at 4.4 Å resolution), EMD-8816 (map after signal-subtracted classification for density near XPD), and EMD-8817 (map after signal-subtracted classification for the XPD 4FeS domain). The atomic coordinate model has been deposited in the Protein Data Bank (PDB) with accession code 5OF4.

Figure Legends

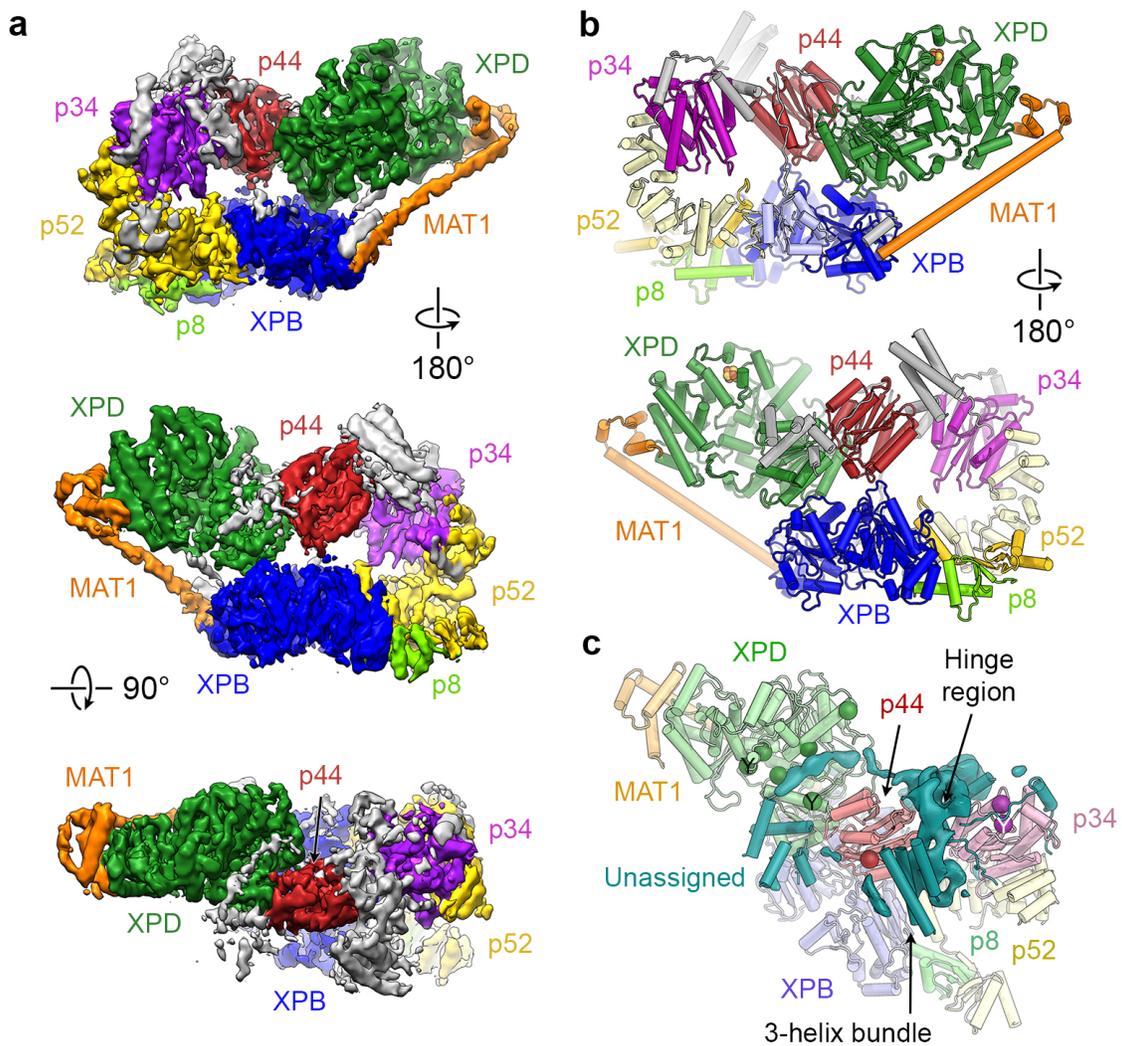


Figure 1 | Cryo-EM reconstruction of human TFIID. **a**, Cryo-EM map of TFIID, color coded and labeled according to constituent subunits; unassigned density grey. **b**, Front and back views of the molecular structure of TFIID. Protein subunits are labeled. Unassigned secondary structure elements attributed to XPB and p52 light blue and pale yellow, respectively; remaining unassigned elements grey. **c**, Unassigned secondary structure elements and remaining unassigned density in the p34-p44 hinge region (teal) may correspond to p62 and the zinc binding domains of p34 and p44. Positions of crosslinks between p62 and other TFIID core proteins ⁷

shown by spheres colored according to the crosslinked partner; crosslinks from studies of yeast TFIID indicated by a Y.

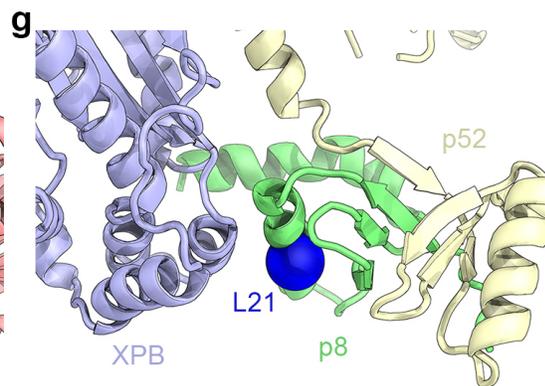
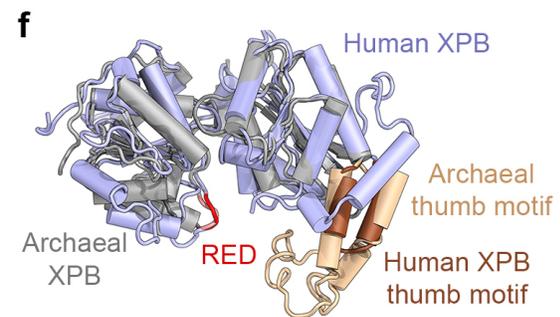
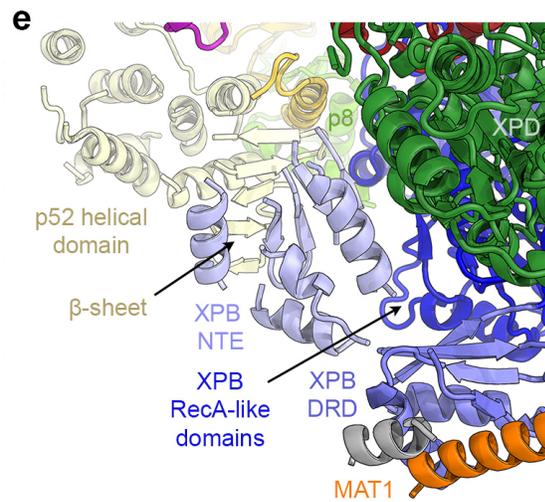
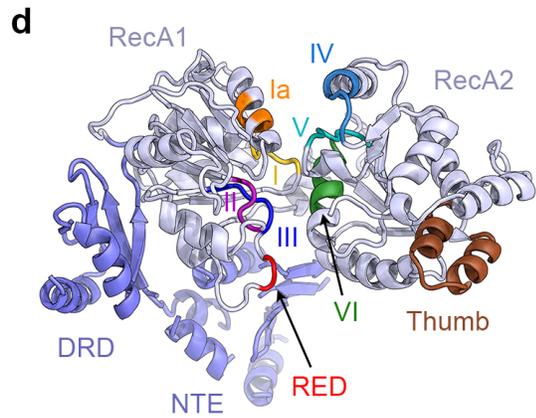
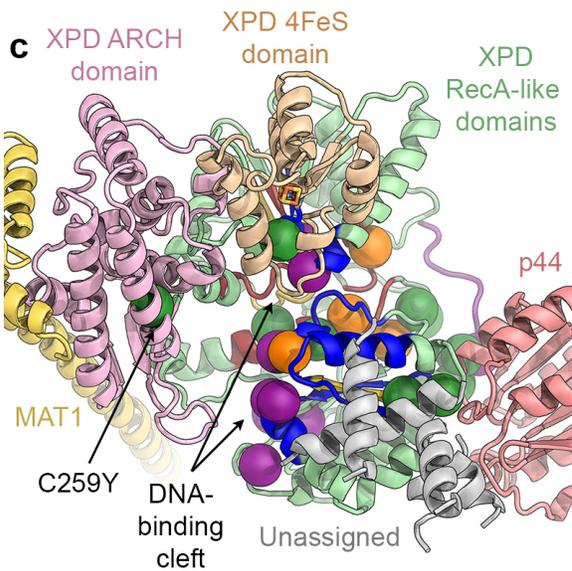
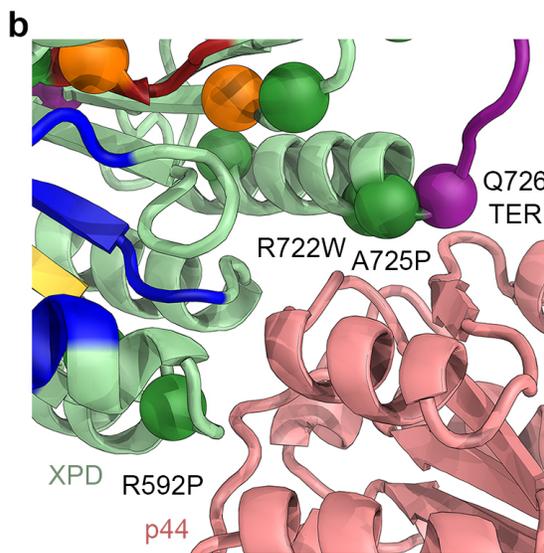
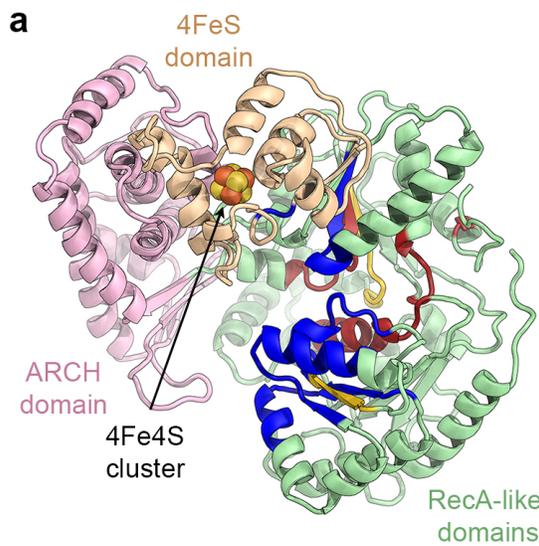


Figure 2 | Structures of human XPB and XPD. **a**, Color-coded domain architecture of XPD: ARCH domain pink, 4FeS domain brown, RecA-like domains green, with helicase motifs distinctively colored by function (dark red: nucleotide binding and hydrolysis; yellow: coordination between nucleotide hydrolysis and helicase activity; blue: DNA binding) ^{10,16}. **b**, Disease-causing mutations mapped onto the structure of human XPD as solid spheres: XP mutations purple, TTD mutations dark green, mutations causing combined XP and Cockayne's syndrome orange. XPD mutations near p44 (light red) are labeled and numbered. Q726 is affected by a nonsense mutation in XP; the protein segment truncated thereby is shown in purple. **c**, XPD in the context of surrounding proteins, with disease-causing mutations mapped onto the structure as in **b**. **d**, Structure of XPB (RecA-like domains light violet, NTE and DRD blue, RED motif red, thumb motif brown, conserved helicase motifs color coded and labeled) ²⁰. **e**, Secondary structure elements modeled into unassigned density (red in Extended Data Fig. 5b) include the XPB NTE and DRD, and a β -sheet tentatively assigned to p52. **f**, Superposition of the RecA-like domains of human (violet) and archaeal (grey) XPB (PDB ID 2FWR ²⁰); domains superposed separately due to a large conformational differences. **g**, p8 (light green) bridges the p52 C-terminal domain (pale yellow) with XPB (violet). The L21P mutation (blue sphere) causes TTD.

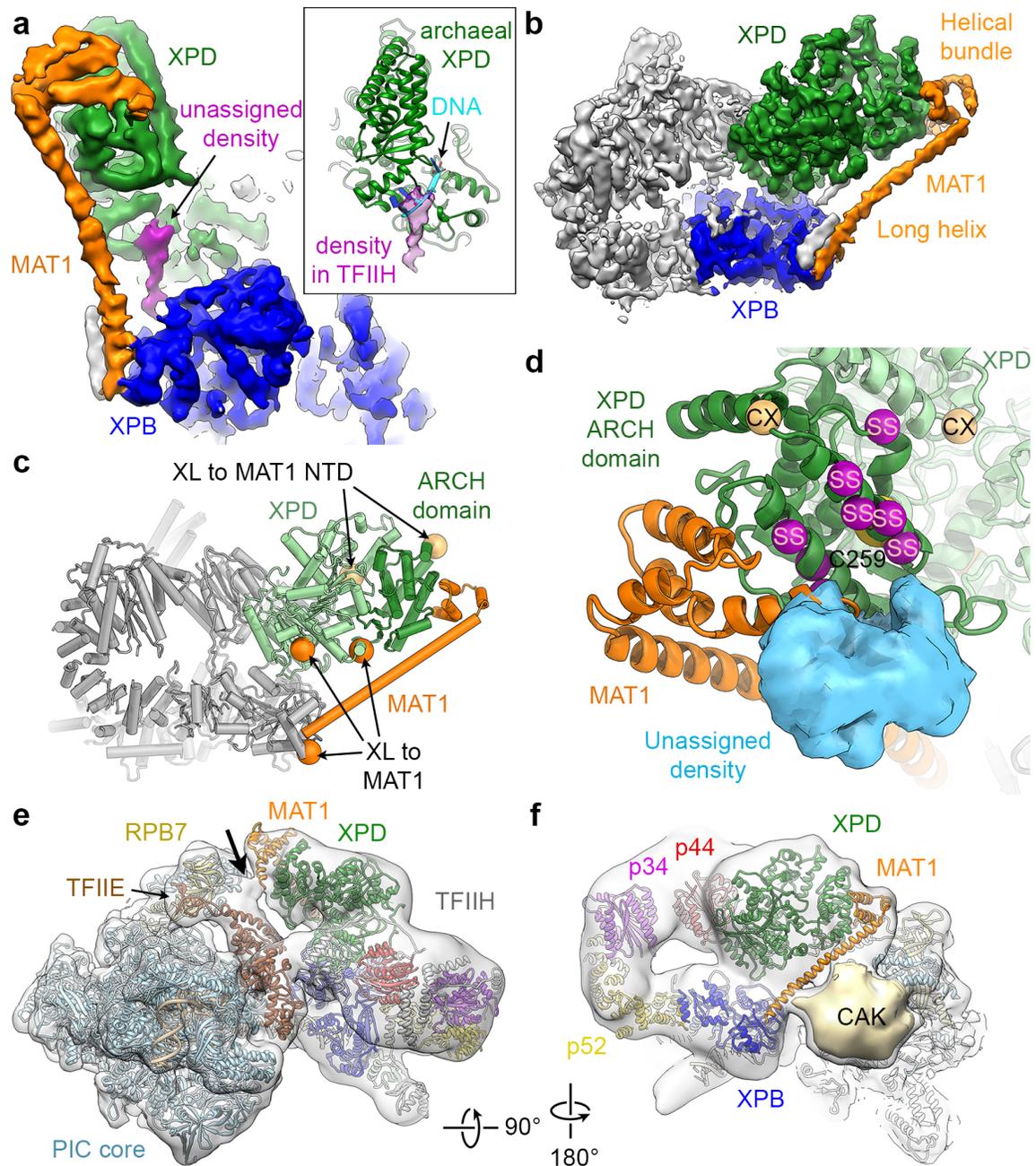


Figure 3 | MAT1 forms a physical connection between all ATP consuming TFIIF moieties. **a, b**, A long α -helical density and a helical bundle (orange) connect XPB (blue) and XPD (green). Unassigned density (purple in **a**) connects the XPB DRD with the DNA-binding site of XPD. Inset: Unassigned density superimposed on the structure of archaeal XPD¹¹ (PDB ID 5H8W; green) bound to DNA (cyan). **c**, Locations of CX-MS crosslinks (spheres)⁷ of human MAT1 (orange) and the N-terminal domain of the yeast MAT1 homolog (light orange) on XPD (green; ARCH

domain dark green). **d**, Interaction region between MAT1 and the XPD ARCH domain. Mutations of C259 (partially occluded yellow sphere) impair the MAT1-XPD interaction. Site-specific crosslinks (SS) between RAD3 and TFB3 (yeast homologs of XPD and MAT1) ²⁹ shown as purple spheres, CX-MS crosslinks (CX; same as in **c**) light orange. Unassigned density that may correspond to the N-terminal RING domain of MAT1 (Extended Data Fig. 7d, e) shown as blue surface. **e**, Docking of the structure of TFIIH into the density of the Pol II-PIC (EMD-8134) ⁵. Density for the MAT1 helical bundle is visible near a contact site (arrow) of TFIIH with the PIC core near TFIIE and RPB7. **f**, As **e**, but viewing TFIIH from the side. Density for the long MAT1 α -helix is not clear, possibly due to flexibility or structural rearrangement. A large but relatively weak density contacting Pol II likely corresponds to the Cyclin H and Cdk7 subunits of the CAK ^{5,6} (Extended Data Fig. 7f, g).

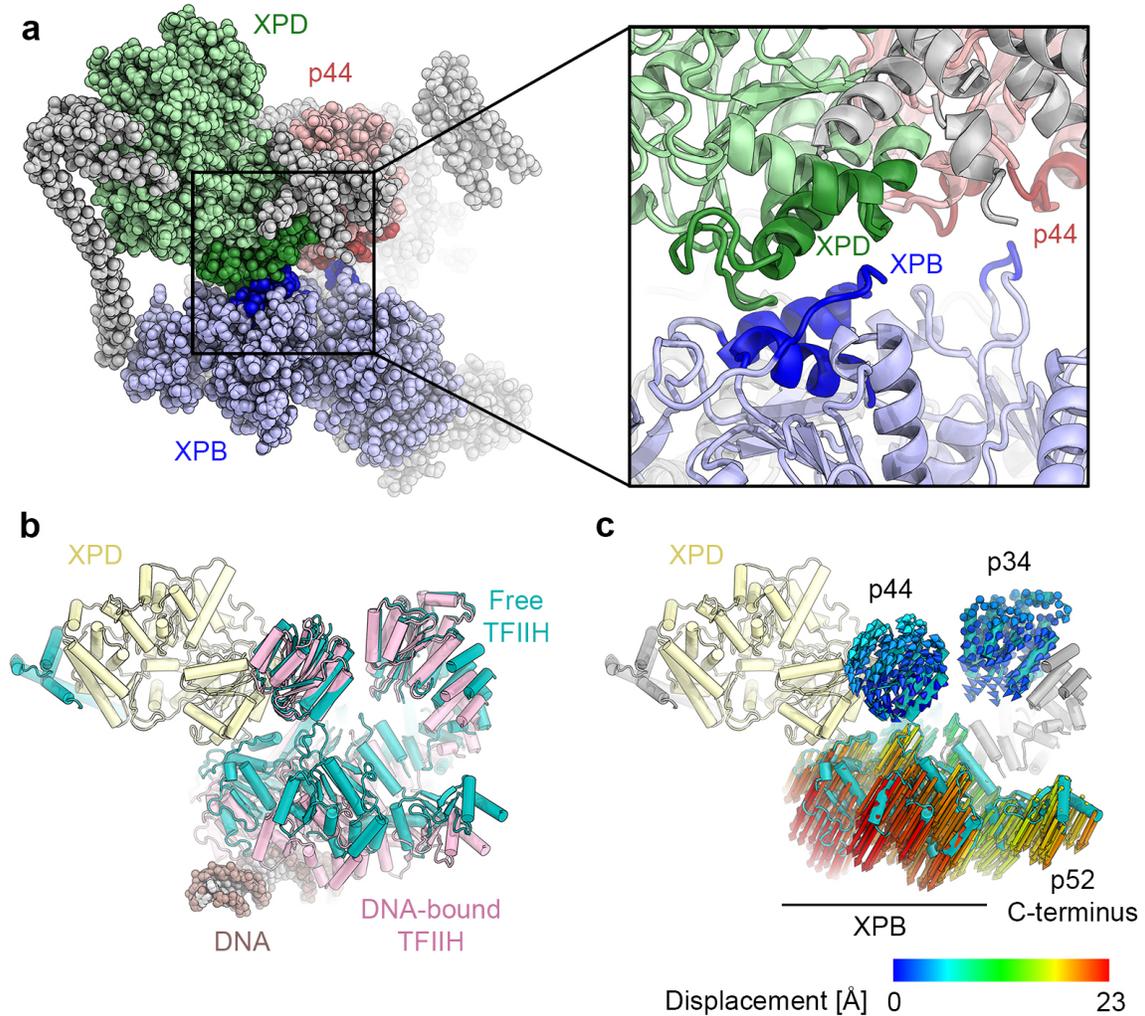


Figure 4 | Conformational rearrangements of TFIIH. **a**, XPD-XPB and p44-XPB contacts in free TFIIH. Interacting regions colored dark blue on XPB, dark green on XPD, and dark red on p44. **b**, Free TFIIH (teal) and TFIIH bound to DNA in the Pol II-PIC (pink; subunits of free TFIIH were fitted into the cryo-EM map of TFIIH in the PIC⁵) superimposed on XPD (light yellow). A conformational change in TFIIH separates the XPB and XPD ATPases when XPB is bound to DNA. **c**, Analysis of XPB and XPD C α displacement during the conformational change from free to PIC-bound TFIIH shows a downward movement of XPB, p8, and the p52 C-terminus, and slight rotations of p34 and p44.

Methods

TFIIH purification and cryo-EM specimen preparation

TFIIH was immuno-purified from HeLa cells as described ^{6,33}. Cryo-EM specimens for data collection were prepared using TFIIH at a concentration of ~10 nM, in 20 mM HEPES-KOH pH 8, 150 mM KCl, 5 mM MgCl₂, 0.5 mM DTT, 2 % trehalose, 1.5 % glycerol and 0.015% NP-40 substitute. 4 µl of sample were applied to freshly plasma cleaned (Solaris, Gatan) C-flat CF 4/2 holey carbon grids (Protochips) coated with a thin film of continuous carbon. After 3-4 min incubation at 100% humidity and 5°C, the grids were blotted for 2 s in a Vitrobot Mark IV (FEI Company) and plunge-frozen in liquid ethane at liquid N₂ temperature.

Cryo-EM data collection and processing

Cryo-EM data was collected on a FEI low-base Titan transmission electron microscope with a Gatan side-entry holder. The microscope was operated at 300 kV acceleration voltage and was equipped with a Gatan K2 Summit direct electron detector camera operated in counting mode. Data was collected semi-automatically using the LEGION package ³⁴ at 37'879 x magnification, resulting in a pixel size of 1.32 Å, with a total electron dose of 40 electrons per Å² during a total exposure time of 8.7 s, fractionated into 30 movie frames. Defocus values ranged from -1.5 µm to -5.0 µm. Four datasets were collected and initially processed separately. Approx. 5800 micrographs from a total of roughly 8300 micrographs were selected for further processing (dataset 1: 1000/1200; dataset 2: 1200/1700; dataset 3: 1800/2800; dataset 4: 1800/2600 micrographs).

Stacks of dose-fractionated image frames were aligned using DOSEFGPU_ DRIFTCORR ³⁵. The drift corrected micrographs and their power spectra were inspected using the APPION pipeline ³⁶. Micrographs with poor power spectrum or excessive ice contamination were removed. Contrast transfer function (CTF)

parameters were estimated using CTFFIND4³⁷ in RELION³⁸. Initially, particles were picked using DOG picker³⁹ in APPION³⁶. After 2D classification, representative classes were used to generate picking reference for RELION autopicking⁴⁰, resulting in a total of 1,513,200 particle picks (dataset 1: 195,500; dataset 2: 258,800; dataset 3: 456,400; dataset 4: 602,500). These picks were initially subjected to 3D and 2D classification (Extended Data Fig. 1a, b) in RELION 1.4 (ref.³⁸). 3D classification was performed before 2D classification because the reconstructions resulting from the use of particles selected with this strategy were of higher quality, possibly due to the retention of more of the rarer views of the particle. During the initial 3D classification, one class consistently showed non-particle features (Extended Data Fig. 1a), which is to be expected during the first overall classification step due to the absence of previous 2D classification. The particles selected during these steps were refined to higher resolution in RELION 1.4 (ref.³⁸) (Extended Data Fig. 1a). At this point, the particles were corrected for local motion and dose weighed using the particle polishing procedure in RELION⁴¹. The polished particles of datasets 1 and 2 were refined and subclassified together, while datasets 3 and 4 were refined and subclassified individually. The 122,900 particles from datasets 2, 3, and 4 selected during these classifications at high resolution were then combined, split into random half-sets (gold standard refinement^{38,42}), and refined to a 4.4 Å cryo-EM map of the TFIIH core complex. Dataset 1 was discarded because the particles it contained did not contribute to high resolution (i.e. β-strand separation was never observed for reconstructions from this dataset alone, Extended Data Fig. 1c). Because local resolution estimation indicates that information significantly beyond 4.4 Å is present in the high-resolution map (Extended Data Fig. 2b), it was sharpened using B-factors of -200 to -300 and low-pass filtered at 4.4 to 4.0 Å resolution for visualization of peripheral parts and well-resolved core regions, respectively.

For signal-subtracted classifications⁴³, the region of interest was extracted from the reconstruction obtained using the final 122,900-particle dataset using a soft mask. The remainder of the complex was subtracted from the particle images as described

⁴³, and the particle images were then used for alignment-free 3D classification into 6 classes. For the XPD/MAT1 region (Extended Data Fig. 2c) and the XPD 4FeS-domain (Extended Data Fig. 2d), the non-subtracted particles corresponding to the signal-subtracted particles assigned to the classes of interest during alignment-free classification were subsequently refined to obtain a map of the entire TFIIH complex. This procedure was described before ^{43,44} and was necessary because the regions of interest on their own were too small to provide sufficient signal for alignment. Likewise, refinement of the entire complex with XPB subtracted, which was attempted because of opening-closing motions of the two halves of the horseshoe-shaped TFIIH assembly that may be limiting the resolution of the reconstructions, did not result in improved resolution, likely because the remaining signal for alignment was insufficient and the alignment quality deteriorated.

Negative stain EM analysis

For preparation of negatively stained TFIIH specimens, 4 μ l of \sim 10 nM of TFIIH in 20 mM Hepes-KOH pH 8.2, 100 mM KCl, 5 mM MgCl₂, 0.5 mM DTT, 3 % trehalose, 5 % glycerol and 0.01 % NP-40 substitute were incubated on plasma-cleaned 400-mesh copper grids coated with a carbon film (supported by nitrocellulose) for 90 sec and then stained in 2 % (w/v) uranyl formate by dipping the grid onto five droplets of stain for 10 sec each, followed by blotting with filter paper.

Images were acquired using a Tecnai F20 transmission electron microscope (FEI Company) operated at 120 kV acceleration voltage equipped with a US4000 CCD camera (Gatan), using defocus values of -0.8 to -1.5 μ m at 107'142 x magnification. Data was collected semi-automatically using the LEGINON software package ³⁴ and initially screened using the APPION pipeline ³⁶ to eliminate micrographs exhibiting excessive drift or poor staining. Approx. 69,500 particles were selected using DOG PICKER ³⁹ from within APPION and then extracted using a box size of 256 pixels at a pixel size of 1.4 \AA /pixel. These images were further processed using RELION ³⁸ after two-fold binning to obtain reference-free 2D class averages. 2D classification was

performed with a larger particle diameter setting of 340 Å to account for the full size of the CAK subcomplex. Two of these classes with well-defined density for the CAK subcomplex are shown in Extended Data Fig. 8a.

For 3D reconstruction of TFIID complexes with CAK density, a separate dataset of 72,500 particle images, which were less strongly preferentially oriented, was used. The conditions for data collection in initial processing were as described above. The particle images were extracted in 256 pixel boxes at 1.4 Å/pixel, four-fold binned, and subjected to 2D classification. From the resulting classes, approx. 50,700 particles were selected and classified into four 3D-classes (Extended Data Fig. 8c), using as an initial 3D reference the density of TFIID in the context of the Pol II-PIC⁵ (EMD-8131).

Molecular modeling and atomic model building

At the the 4.4 Å overall resolution obtained for the reconstruction of the TFIID core complex, likely limited by the flexibility of the complex, *de novo* chain tracing and register assignment was not possible in many regions of the density. Therefore, our modeling relied to a large degree on docking of homology models, their rebuilding, and extension of their termini. Additionally, we built idealized secondary structure elements into unassigned density. For most of the model building, the final map at 4.4 Å resolution was used, except for parts of the 4FeS-domain of XPD, where the map resulting from signal-subtracted classification (Extended Data Fig. 2d) helped guide the model building. For the proteins that we could assign in our density, a combination of density for aromatic side chains and homology to existing high-resolution structures guided assignment of the sequence register during rebuilding. However, at the present resolution, register shifts cannot be excluded in the less-well resolved parts of the map.

Homology models for the human variants of XPD, the N-terminal ATPase domain of XPB, p44, p34, and the complex between p8 and the C-terminal domain of p52 were

created using PHYRE2⁴⁵ based on existing high-resolution structures. These homology models and the X-ray crystal structure of the human XPB C-terminal ATPase domain (PDB ID 4ERN⁴⁶) provided the starting point for model building. All models were rebuilt using O⁴⁷ and COOT⁴⁸ according to the cryo-EM map. The XPD homology model (based on PDB ID 2VSF⁸ and complemented by PDB ID 3CRV¹⁰) was extensively rebuilt where discrepancies between the archaeal template and the density were apparent. The model was extended at its N- and C-termini, and the insertion element in the ARCH domain (residues 288-318) was built into the density. In the 4FeS-domain, which needed to be adjusted extensively because of structural differences with the archaeal template, the Fe₄S₄ cluster was visible as a strong density peak and placed accordingly (Extended Data Fig. 6a, b). For XPB, the homology model of the N-terminal RecA-like domain (based on PDB ID 2FWR²⁰) and the crystal structure of the C-terminal RecA-like domain (PDB ID 4ERN⁴⁶) were rebuilt according to the density. The structures of the archaeal DRD (PDB ID 2FWR²⁰) and the SMARCAL1 HARP1 domain (PDB ID 4O66⁴⁹) were used to guide the modeling of the DRD, based on the results of an unbiased molecular replacement search of the density in this region (see below). The HARP1 domain crystallized as a domain-swapped dimer⁴⁹ but was resolved into a single domain for modeling. Due to the pseudo-symmetry of the DRD and HARP1 domain, two orientations related by a 180° rotation are compatible with the map. However, only one of these orientations places the termini of the docked domains near the connecting density to the N-terminal RecA-like domain and the NTE domain of XPB. Therefore, this orientation (shown in Extended Data Fig. 7) was chosen for modeling. Residues 222-256 were modeled as a poly-alanine chain and deposited as UNK because a disordered segment with no visible density interrupted the trace and precluded register assignment in this region. The models of p34 and p44 (based on PDB IDs 4PN7⁵⁰ and 4WFQ⁵¹, respectively) were rebuilt similarly. Because the density for the p52 C-terminal domain and p8 (modeling based on the yeast p8/p52 dimer, PDB ID 3DOM²⁵, and the structure of human p8, PDB ID 2JNJ⁵²) was less-well resolved than the remainder of the structure, adjustments were mostly limited to movement of rigid bodies, or entire α -helices and β -sheets.

After docking and adjustment of these structures and models, significant amounts of cryo-EM density remained unassigned. Idealized secondary structure elements, both α -helices and β -strands, were built into this density using O⁴⁷ and COOT⁴⁸. The cryo-EM map also revealed a number of extended linkers that run across the surface of the TFIIH core complex. These were modeled as extended amino acid chains. All of these structural elements were modeled as poly-alanine chains and deposited in the final model as UNK. These newly modeled secondary structure elements were divided into four groups (Extended Data Table 1): MAT1 (chain H), and chains X, Y, and Z. Even though definitive assignment of these chains would require a map that allows full chain tracing and register assignment, existing experimental data provides sufficient evidence for a tentative assignment. MAT1 was assigned as discussed in the text (Fig. 3, Extended Data Fig. 8) based on secondary structure prediction (Extended Data Fig. 4)⁴⁵, crosslinking^{7,29}, and biochemical and mutation data²⁸. Chain X encompasses several α -helices and a β -sheet near the RecA-like domains of XPB and likely represents the NTE domain of XPB, which is specific to eukaryotes and not present in the archaeal X-ray crystal structures. Chain Y encompasses a set of α -helices that connect XPB to p34. Based on biochemical results, secondary structure prediction, and the location of the p52 C-terminal domain, these helices as well as a β -sheet in their immediate vicinity most likely form part of p52. The remaining elements are grouped in chain Z, some of which may form part of p62, based on CX-MS data⁷ (see text and Supplementary Discussion).

Molecular replacement search

An unbiased molecular replacement search using the BALBES-MOLREP pipeline⁵³⁻⁵⁵ was performed for unassigned regions in the TFIIH cryo-EM map. The search included the density eventually assigned to the DRD and the NTE of XPB (red in Extended Data Fig. 5b), as well as the unassigned densities in the vicinity of p34 (teal in Extended Data Fig. 5a, b, d). Significant hits, as judged by the contrast score

and the fit of the model to the map were only found for the region subsequently assigned to the DRD. A fragment of PDB ID 2BOP, depicted in Extended Data Fig. 7a-c, was the hit with the second highest contrast score (6.03; contrast score > 3.0 is indicative of a correct solution⁵³); the best hit was a fragment from PDB 1W9K (contrast score 7.03), which shows a similar fold overall, but is less complete (not shown).

Coordinate refinement

The atomic coordinate model was initially refined in PHENIX⁵⁶ using the real space refinement algorithm implemented in the program `phenix.real_space_refine` into a masked map that had previously been sharpened by a B-factor of -300 \AA^2 and filtered to 4 \AA resolution; however, only data to the overall 4.4 \AA resolution was used for refinement. The model was subjected to 5 macro-cycles of rigid body refinement in real space using single chains as rigid body groups to ensure proper fit of the model to the map used for refinement, followed by 6 macro cycles of global optimization in real space. Real space global minimization was performed using Ramachandran, rotamer, C_β , and secondary structure restraints to maintain model geometry. Secondary structure restraints were automatically generated in PHENIX using the `ksdssp` algorithm (based on ref.⁵⁷) and manually inspected and adjusted. Custom bond lengths and bond angles were used to stabilize the coordination geometry of the $4\text{Fe}4\text{S}$ cluster. After the first round of refinement, the coordinate model was inspected and errors were corrected. Subsequently, the model was refined again for 6 cycles using PHENIX real space refinement as described above. The final refinement was performed in REFMAC5⁵⁸ in reciprocal space as described previously⁵⁹ and using secondary structure restraints generated by PROSMART⁶⁰. The refinement was performed against structure factor amplitudes and phases computed from a map that had been masked, sharpened (B-factor of -250 \AA^2), and filtered (to 4.2 \AA resolution). To avoid over-fitting, only spatial frequencies of 4.4 \AA or lower were used in the refinement. A refinement weight of 0.001 was applied, and the geometry of the molecular model was additionally stabilized by applying

“jelly-body” restraints⁵⁸. The model was refined for 20 cycles, resulting in an R-factor of 33.7 %, which may be somewhat lower than for a comparable X-ray refinement due to the relatively larger unit cell and higher solvent content in the cryo-EM map.

The FSC curve between the refined model and the 4.4 Å cryo-EM map of the TFIID core complex (Extended Data Fig. 2a) shows that meaningful information is present beyond the nominal 4.4 Å resolution of the cryo-EM density. Refinement statistics, model validation and the distribution of B-factors on the refined coordinate model are shown in Extended Data Fig. 3g-i. To monitor over-fitting, we validated the coordinate model using half-map refinement as described previously⁵⁹. The absence of a large gap between FSC_{work} and FSC_{test} indicates that the refinement weight chosen during coordinate refinement did not induce excessive overfitting (Extended Data Fig. 3i).

Creation of figures

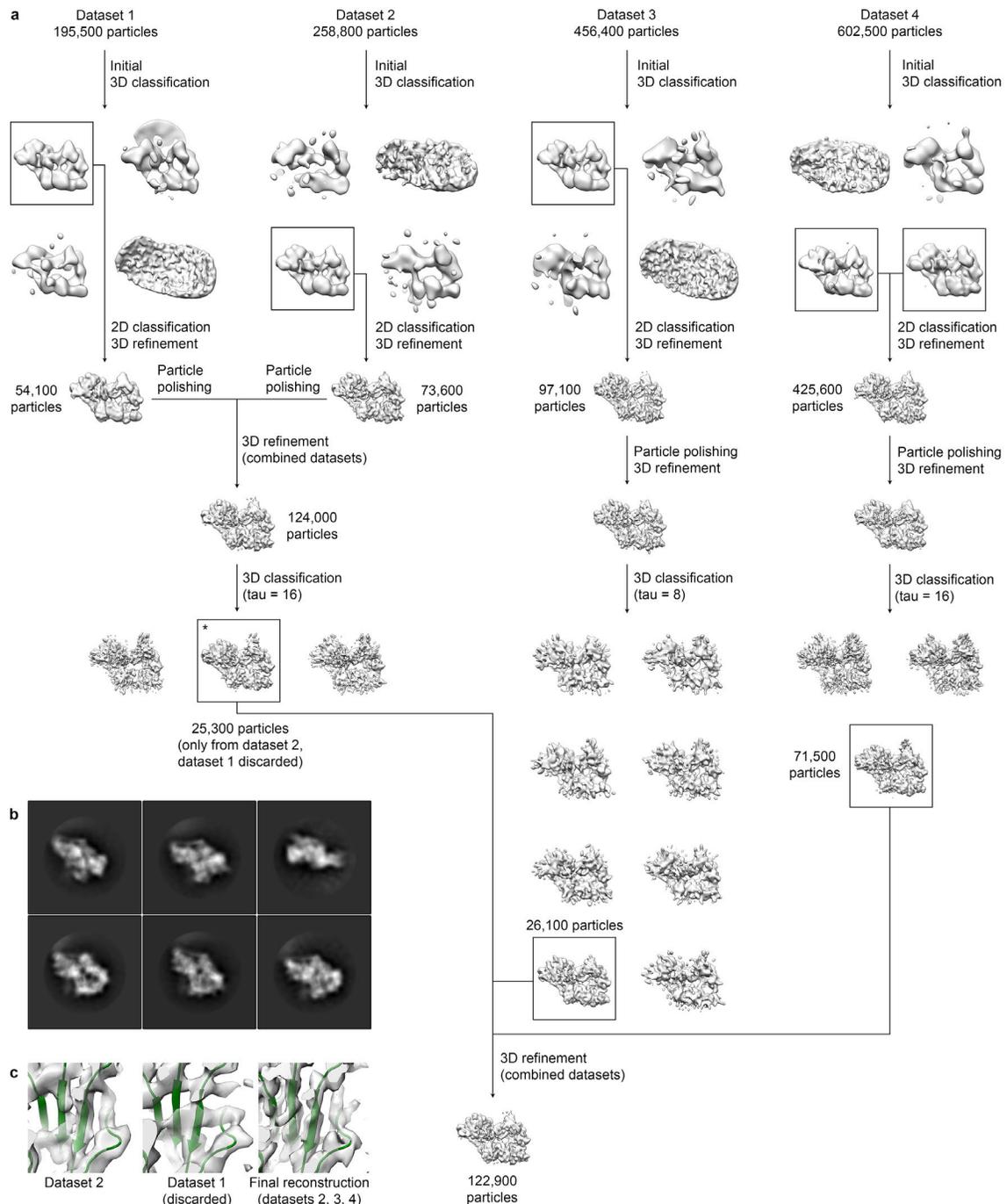
Figures were created using PyMol (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.) and the UCSF Chimera package⁶¹ from the Computer Graphics Laboratory, University of California, San Francisco (supported by NIH P41 RR-01081).

- 33 Knuesel, M. T., Meyer, K. D., Bernecky, C. & Taatjes, D. J. The human CDK8 subcomplex is a molecular switch that controls Mediator coactivator function. *Genes Dev.* **23**, 439-451 (2009).
- 34 Suloway, C. *et al.* Automated molecular microscopy: the new Legimon system. *J. Struct. Biol.* **151**, 41-60 (2005).
- 35 Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Meth.* **10**, 584-590 (2013).
- 36 Lander, G. C. *et al.* Appion: an integrated, database-driven pipeline to facilitate EM image processing. *J. Struct. Biol.* **166**, 95-102 (2009).
- 37 Rohou, A. & Grigorieff, N. CTFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216-221 (2015).

- 38 Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519-530 (2012).
- 39 Voss, N. R., Yoshioka, C. K., Radermacher, M., Potter, C. S. & Carragher, B. DoG Picker and TiltPicker: software tools to facilitate particle selection in single particle electron microscopy. *J. Struct. Biol.* **166**, 205-213 (2009).
- 40 Scheres, S. H. W. Semi-automated selection of cryo-EM particles in RELION-1.3. *J. Struct. Biol.* **189**, 114-122 (2015).
- 41 Scheres, S. H. W. Beam-induced motion correction for sub-megadalton cryo-EM particles. *Elife* **3**, e03665 (2014).
- 42 Scheres, S. H. W. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nat. Meth.* **9**, 853-854 (2012).
- 43 Bai, X.-C., Rajendra, E., Yang, G., Shi, Y. & Scheres, S. H. Sampling the conformational space of the catalytic subunit of human γ -secretase. *Elife* **4**, e11182 (2015).
- 44 Nguyen, T. H. D. *et al.* Cryo-EM structure of the yeast U4/U6.U5 tri-snRNP at 3.7 Å resolution. *Nature* **530**, 298-302 (2016).
- 45 Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845-858 (2015).
- 46 Hilario, E., Li, Y., Nobumori, Y., Liu, X. & Fan, L. 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 (2013).
- 47 Jones, T. A. Interactive electron-density map interpretation: from INTER to O. *Acta Crystallogr D Biol Crystallogr* **60**, 2115-2125 (2004).
- 48 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501 (2010).
- 49 Mason, A. C. *et al.* A structure-specific nucleic acid-binding domain conserved among DNA repair proteins. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 7618-7623 (2014).
- 50 Schmitt, D. R., Kuper, J., Elias, A. & Kisker, C. The structure of the TFIIH p34 subunit reveals a von Willebrand factor A like fold. *PLoS ONE* **9**, e102389 (2014).
- 51 Kim, J. S. *et al.* Crystal structure of the Rad3/XPD regulatory domain of Ssl1/p44. *J. Biol. Chem.* **290**, 8321-8330 (2015).
- 52 Vitorino, M. *et al.* Solution structure and self-association properties of the p8 TFIIH subunit responsible for trichothiodystrophy. *J. Mol. Biol.* **368**, 473-480 (2007).
- 53 Brown, A. *et al.* Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions. *Acta Crystallogr D Biol Crystallogr* **71**, 136-153 (2015).
- 54 Long, F., Vagin, A. A., Young, P. & Murshudov, G. N. BALBES: a molecular-replacement pipeline. *Acta Crystallogr D Biol Crystallogr* **64**, 125-132 (2008).
- 55 Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr D Biol Crystallogr* **66**, 22-25 (2010).

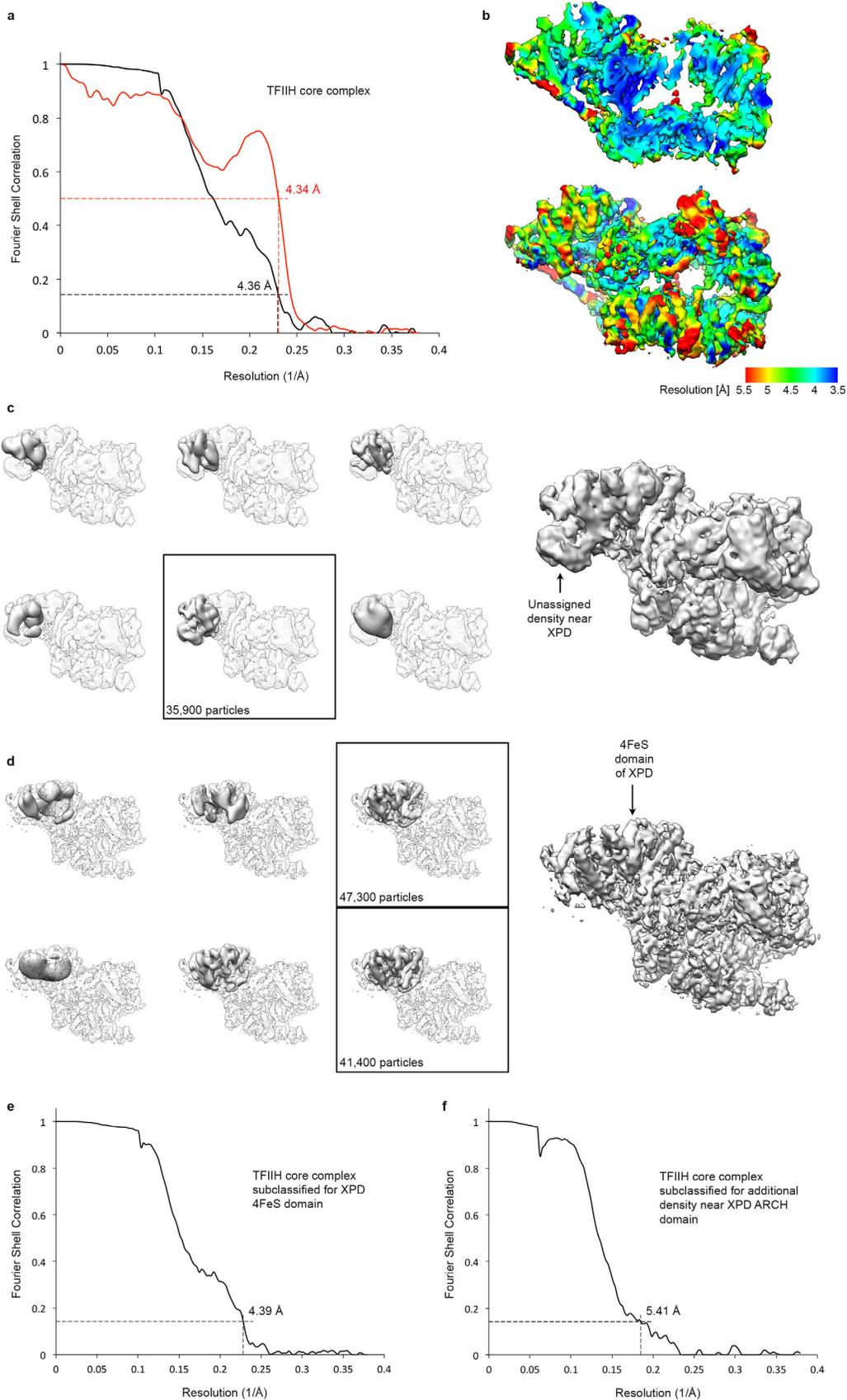
- 56 Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221 (2010).
- 57 Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577-2637 (1983).
- 58 Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* **67**, 355-367 (2011).
- 59 Fernandez, I. S., Bai, X.-C., Murshudov, G., Scheres, S. H. W. & Ramakrishnan, V. Initiation of translation by cricket paralysis virus IRES requires its translocation in the ribosome. *Cell* **157**, 823-831 (2014).
- 60 Nicholls, R. A., Fischer, M., McNicholas, S. & Murshudov, G. N. Conformation-independent structural comparison of macromolecules with ProSMART. *Acta Crystallogr D Biol Crystallogr* **70**, 2487-2499 (2014).
- 61 Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605-1612 (2004).
- 62 Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721-745 (2003).
- 63 Heymann, J. B. & Belnap, D. M. Bsoft: image processing and molecular modeling for electron microscopy. *J. Struct. Biol.* **157**, 3-18 (2007).
- 64 Kim, J. L. *et al.* Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: the crystal structure provides insights into the mode of unwinding. *Structure* **6**, 89-100 (1998).
- 65 Obmolova, G., Ban, C., Hsieh, P. & Yang, W. Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA. *Nature* **407**, 703-710 (2000).
- 66 Gervais, V. *et al.* Solution structure of the N-terminal domain of the human TFIIF MAT1 subunit: new insights into the RING finger family. *J. Biol. Chem.* **276**, 7457-7464 (2001).

Extended Data Figure Legends

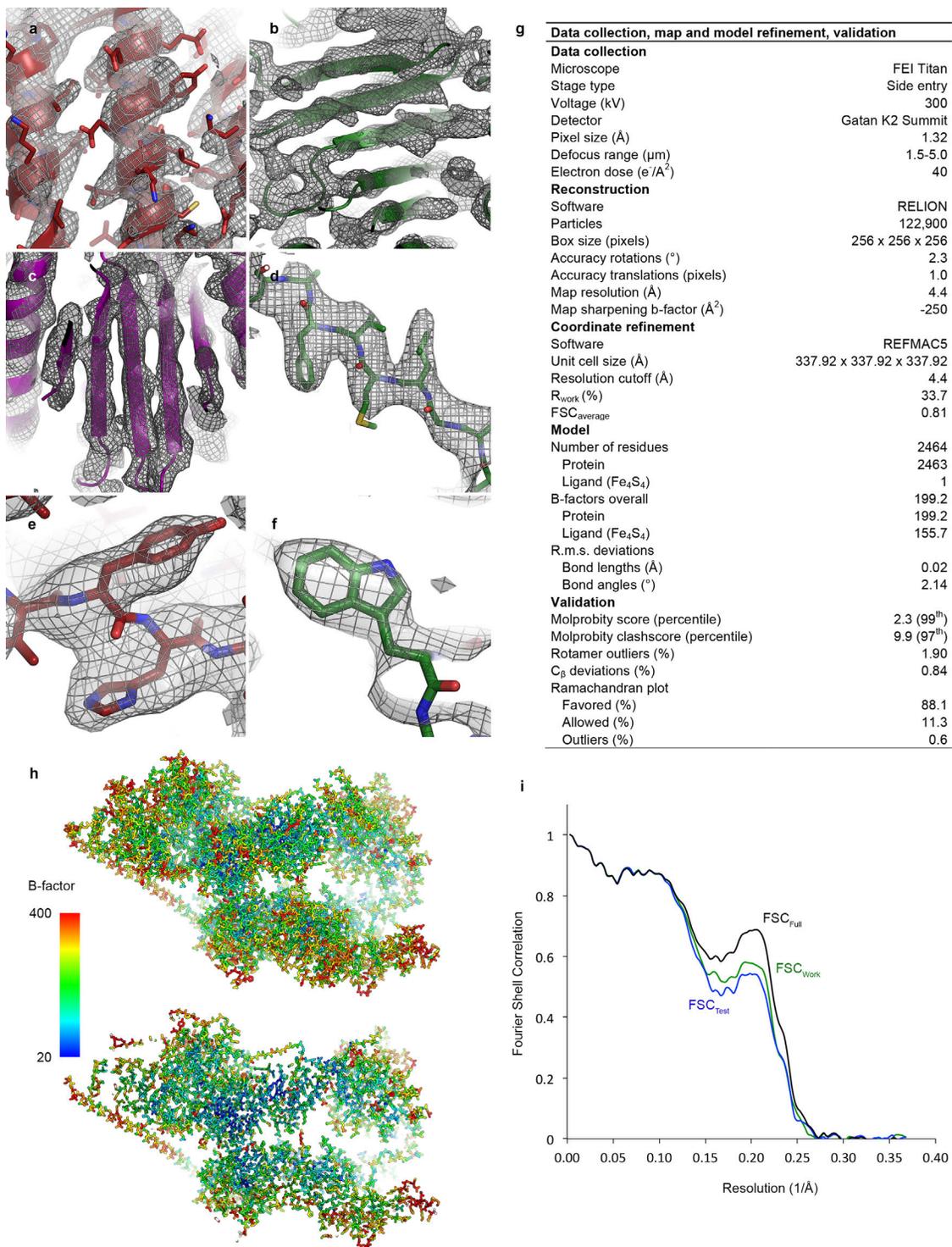


Extended Data Figure 1 | Data processing and cryo-EM map refinement. a, Schematic of data processing and classification procedure. See text and Methods section for details. **b,** Representative class averages obtained during the reference-

free 2D classification step in **a** (taken from dataset 4). **c**, Comparison of the best-resolved densities of intermediate stage classified maps (asterisk in **a**) from dataset 2 (left), dataset 1 (middle), and the final map (shown for comparison, right). The XPD coordinate model is shown in green. Due to the lack of β -strand separation, dataset 1 was discarded from the final refinement (see Methods section for details).

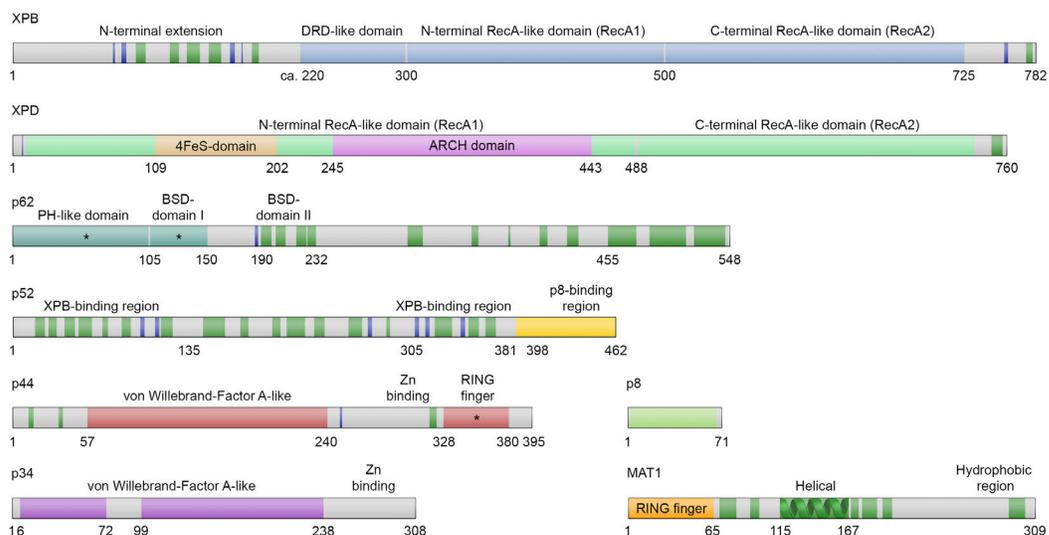


Extended Data Figure 2 | Resolution estimation and focused classification. a, FSC curves. Black: FSC curve for the cryo-EM map of the TFIIH core complex, computed from auto-masked half-maps in RELION³⁸, indicating a resolution of 4.4 Å according to the gold-standard FSC = 0.143 criterion^{42,62}. Red: FSC curve between the cryo-EM map and the refined atomic coordinates. The coordinate refinement used data until 4.4 Å; meaningful correlation between the model and the map extends to 4.3 Å (FSC=0.5 criterion). **b,** Local resolution estimation using the BLOCRES command of the BSOFIT package⁶³. Most of the core of the density of the cryo-EM reconstruction is resolved at 4 Å or better. **c, d,** Signal-subtracted 3D classifications to resolve heterogeneity near the MAT1 binding site on XPD (**c**) and of the XPD 4FeS domain (**d**). The 122,900 particles used to reconstruct the TFIIH cryo-EM map at 4.4 Å resolution were used for this step. 3D classes obtained from signal-subtracted particles are shown on the left, overlaid with a transparent depiction of the entire TFIIH map for orientation. Boxed classes were refined (two classes were combined in **d**). The cryo-EM map obtained by refining the non-signal-subtracted particles corresponding to the particles assigned to the boxed 3D classes is shown on the right. **e, f,** FSC curves for the reconstructions obtained after signal-subtracted 3D classification (**c, d**) for the XPD 4FeS domain (**e**) and the additional density near the XPD ARCH domain (**f**).

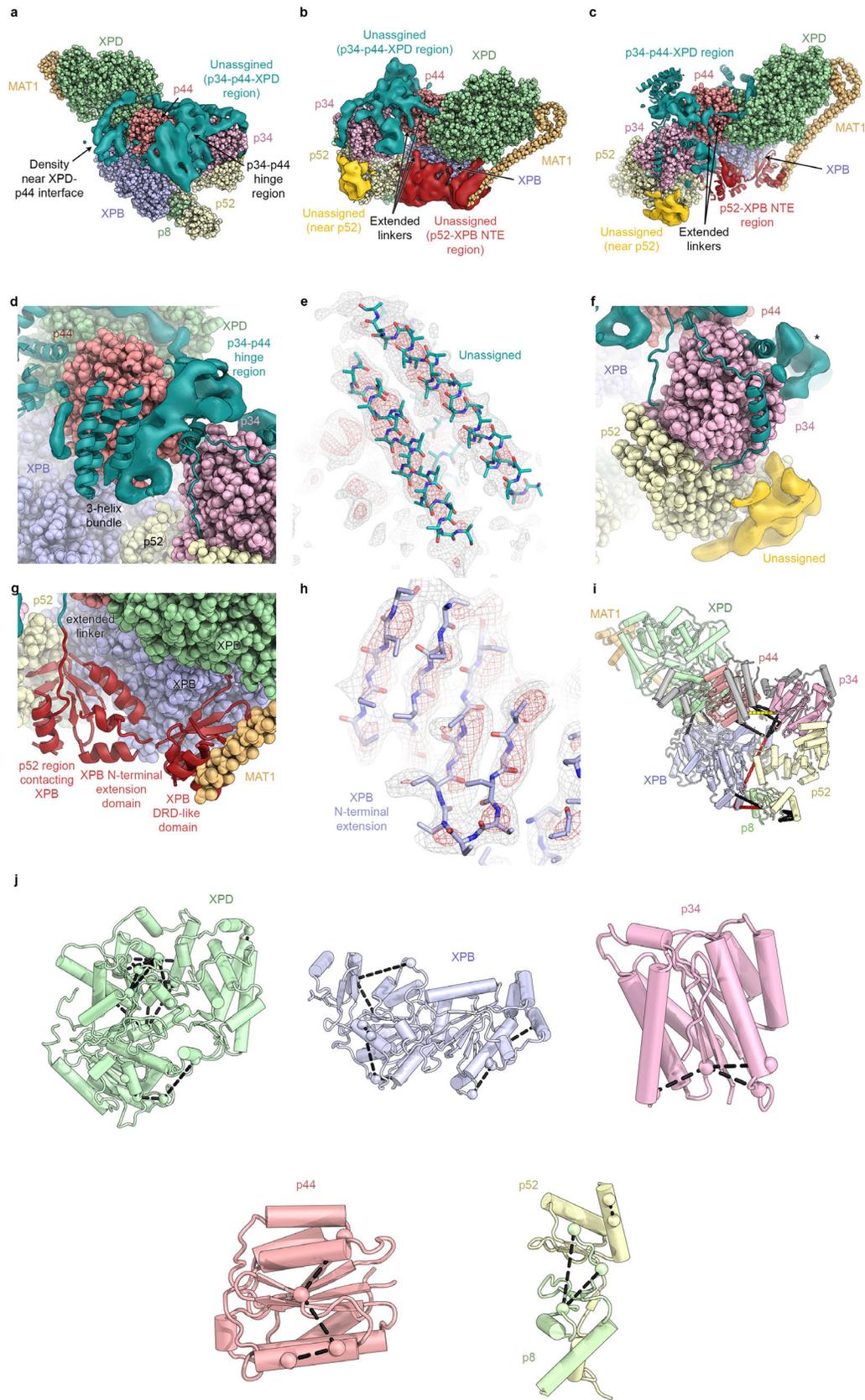


Extended Data Figure 3 | Coordinate refinement and atomic model of TFIID. a-f, Examples of the cryo-EM density of the TFIID core complex, showing α -helices with side chain densities (**a**), separation of β -strands (**b, c**), and density for large side chains (**d-f**). **g**, Refinement table summarizing the key statistics from the atomic

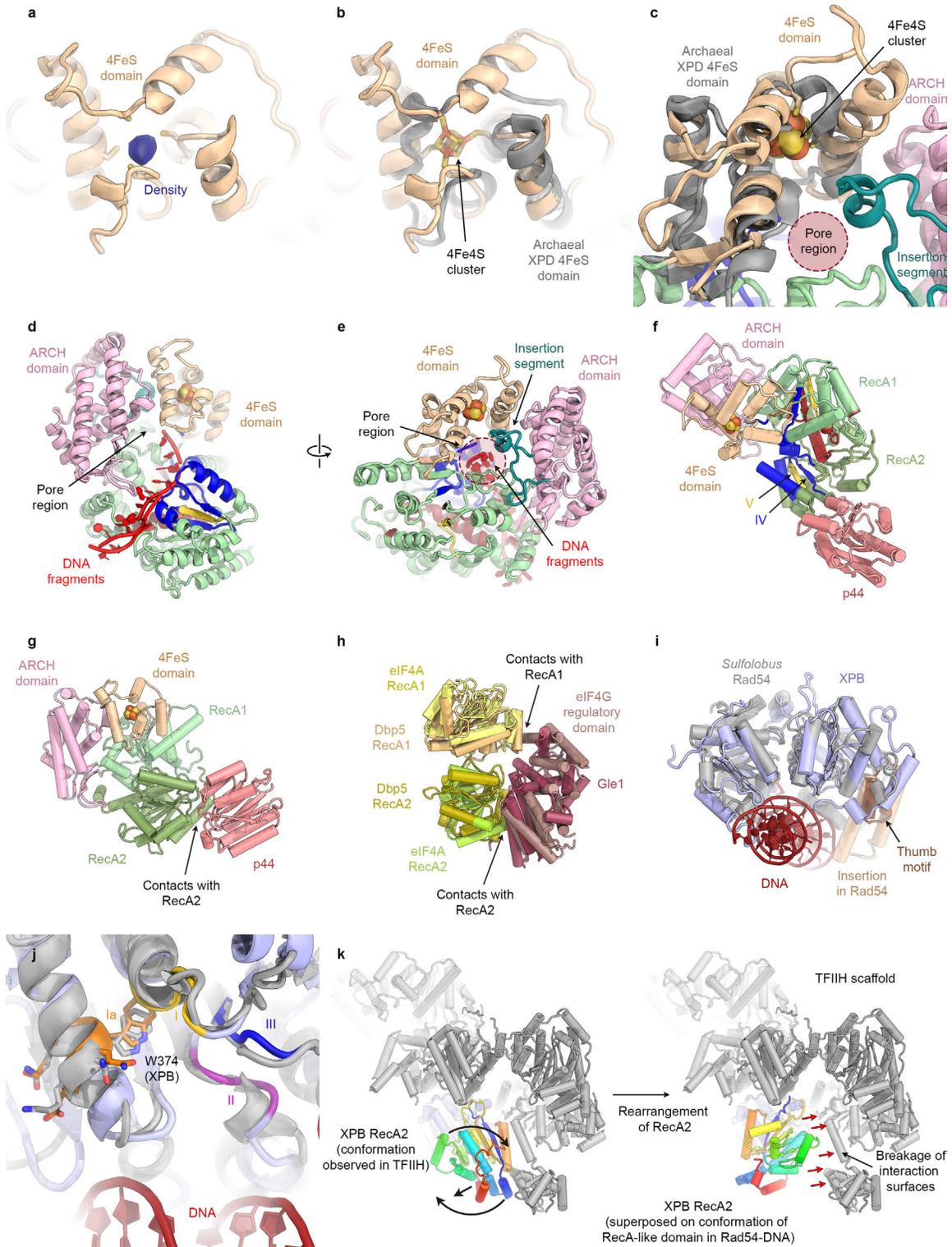
coordinate refinement. **h**, Representation of the refined atomic coordinate model colored according to B-factors (top), and viewed in section (bottom), showing lower B-factors in the better-ordered core of the complex, as expected. **i**, Model validation using data-halfsets. The refined model was refined against one of the gold-standard half maps (work map). The FSC curve (FSC_{work}) between the refined model and the work map is shown in green; the FSC curve (FSC_{test}) between the model and the other half map (test map, not used in refinement) is shown in blue. The half maps were summed and the model was also refined against the summed half maps (black FSC curve, FSC_{full}).



Extended Data Figure 4 | Domain architectures and secondary structure predictions of TFIIH components. Domain organization and secondary structure prediction of the protein subunits of the TFIIH core complex and MAT1. Proteins and domains with known 3D structures (human proteins) are shown as colored areas (colors used correspond to those utilized for molecular depictions) and labeled. All of these are modeled in our molecular structure, except those regions in p44 and p62 marked with an asterisk (*) and the MAT1 ring finger domain. Dark green bars on the grey background indicate predicted α -helices, dark blue bars represent predicted β -strands (predictions were obtained using PHYRE2⁴⁵). Only secondary structure elements predicted with high confidence are displayed. The longest predicted α -helix of the entire TFIIH complex is found in MAT1 and measures more than 50 residues (indicated with a schematic helix).

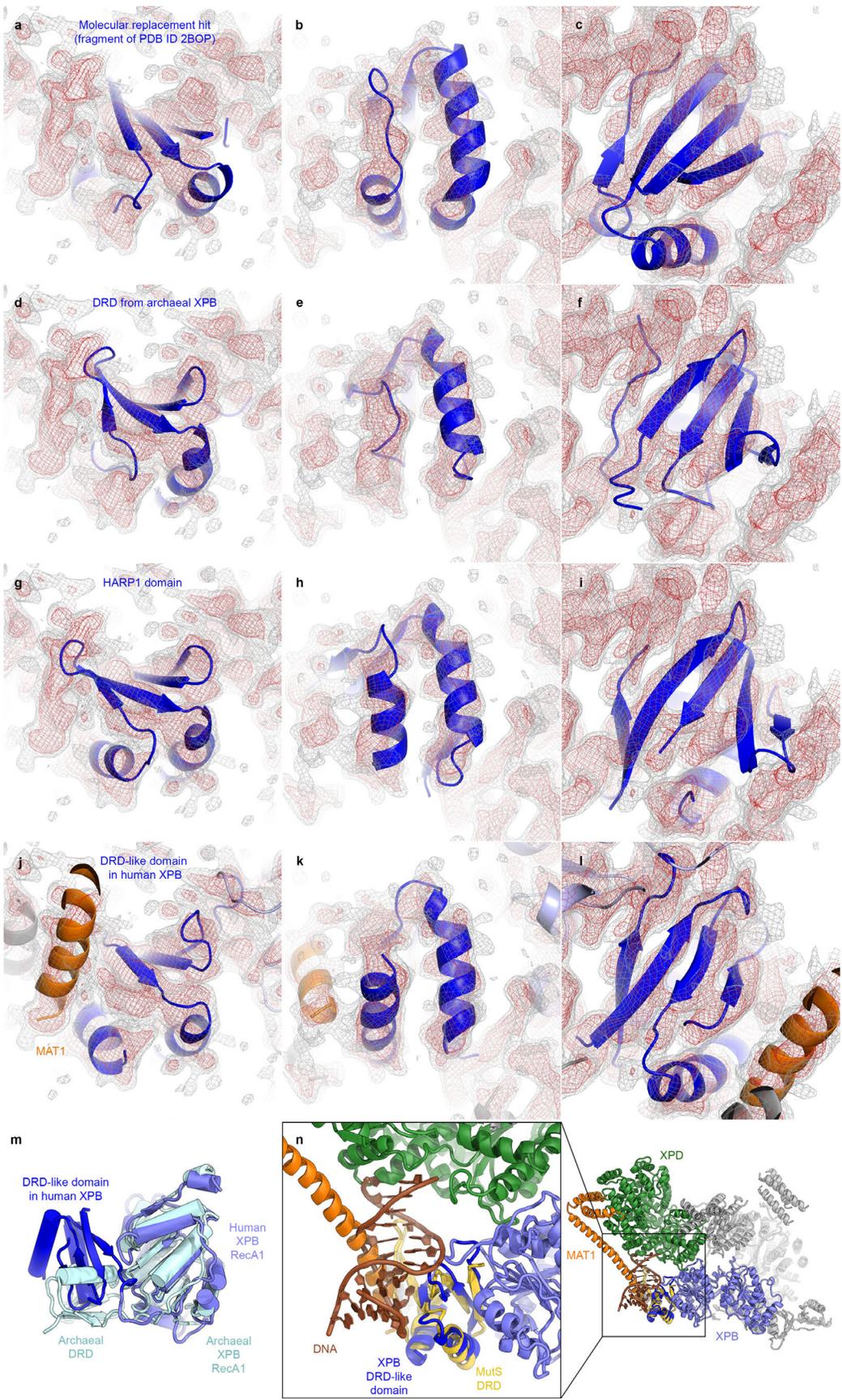


Extended Data Figure 5 | Modeling of secondary structure elements into regions of previously unassigned density. a, b, Unassigned density around the p34-p44-XPB region is shown in teal, around the p52-XPB NTE region in red, and near the p52 helical domain in gold. Densities are low-pass filtered to 6 Å for clarity. Some of these densities could be interpreted by placement of secondary structure elements and extended linkers. **c**, Overview of the molecular model after placement of secondary structures into unassigned densities. Remaining unassigned densities are shown in the same colors as in **a, b** and likely harbor parts of p62 and the zinc-binding domains of p34 and p44. Newly modeled elements are shown in cartoon representation in the same colors as the density regions into which they were modeled. Densities are low-pass filtered to 6 Å for clarity. **d**, Several α -helices could be placed near the hinge region between p34 and p44, though some density remains unassigned (shown as teal surface). **e**, Density and model for the three-helix bundle near the hinge region (**d**). **f**, In addition to the density of the hinge region (**d**), several linkers and α -helices wrap around the p34 vWFA-like domain core, and a small domain (asterisk) is bound to its side. **g**, Secondary structure elements (red) modeled into unassigned density (red density in **b**) in the vicinity of the XPB RecA-like domains and the p52 CTD include the XPB NTE and DRD and a β -sheet tentatively assigned the p52 (see Fig. 2e). **h**, Density and model for one of the β -sheets of the XPB NTE. **i, j**, Mapping of CX-MS crosslinks ⁷ (summarized in Supplementary Data Table 1) onto the structure of TFIIF. Crosslinks are color-coded (Black: C α -C α distance < 30 Å; yellow: 30-35 Å; red: > 35 Å). Intermolecular crosslinks are shown in **i**, intramolecular crosslinks in **j**.

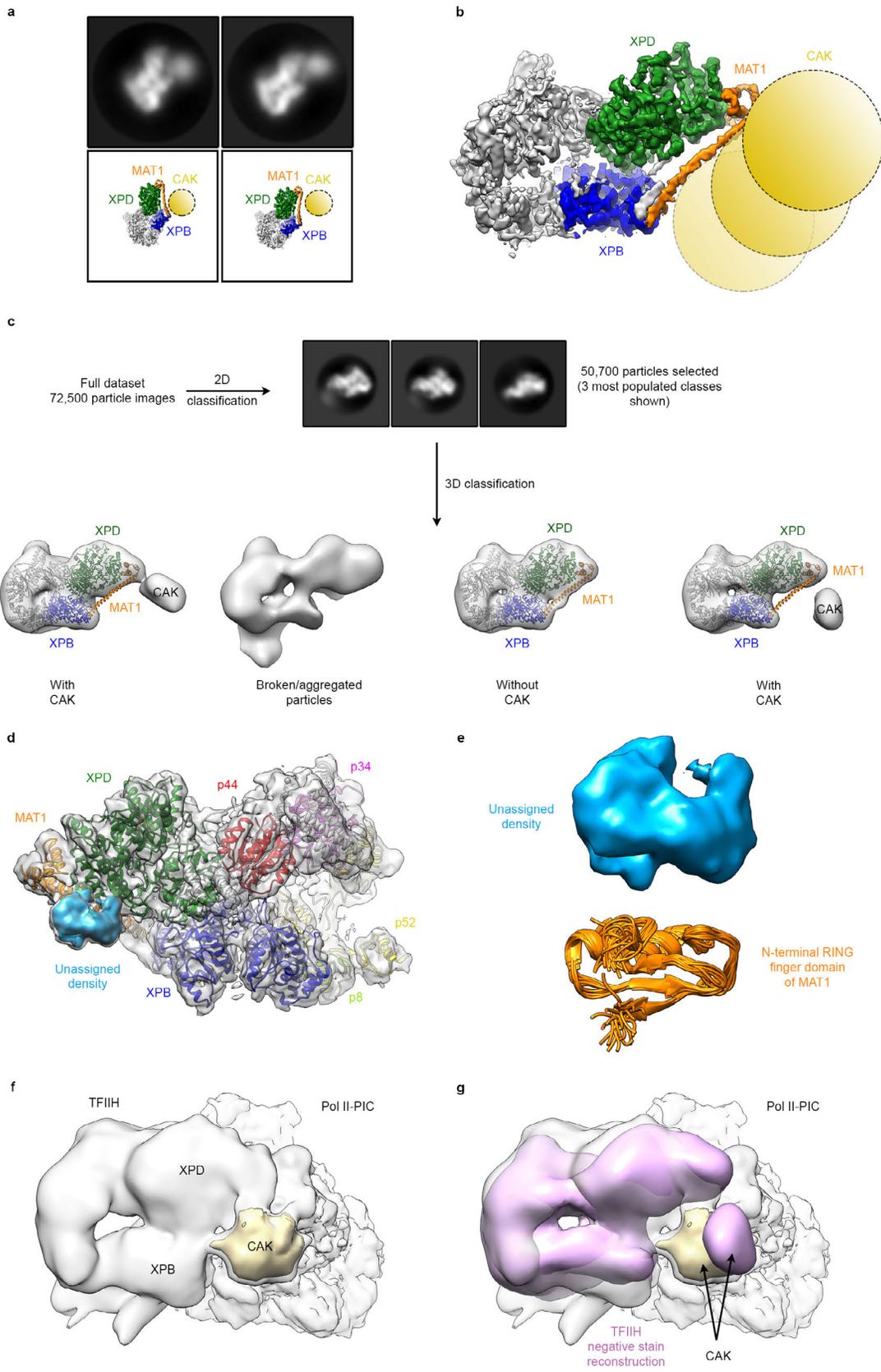


Extended Data Figure 6 | Analysis of the human XPB and XPD structures. a, A cryo-EM density peak (dark blue surface) for the 4Fe4S cluster can be seen within the 4FeS domain of XPD. **b,** Comparison of the structure of the 4FeS domains in

human (brown) and archaeal XPD (grey; PDB ID 3CRV) ¹⁰. The 4Fe4S cluster and coordinating cysteine residues of the human protein are shown in stick representation. **c**, On the back side of XPD, the 4Fe4S-cluster is in proximity of the pore region, which may position it to participate in charge-transfer reactions with DNA passing through the pore ^{8,10}. **d**, Model for the possible path of DNA on XPD obtained by superimposing the DNA-bound structures of archaeal XPD enzymes (PDB ID 4A15, 5H8W) ^{11,12} and the HCV NS3 helicase (PDB ID 1A1V) ⁶⁴. For the NS3 helicase, the two RecA-like domains were superposed individually. The DNA fragments shown in red suggest a possible path of substrate DNA in XPD from the pore between the ARCH and 4FeS domains towards RecA2. The helicase motifs ^{10,16} are colored based on known function (blue: DNA binding, dark red: nucleotide binding and hydrolysis; yellow: coordination between ATPase and helicase activity). **e**, Same as **d**, but viewed from the back side of the pore (as in **c**). **f**, p44 interacts with XPD RecA2 near the conserved helicase elements IV and V (helicase elements indicated; XPD colored as in **d**). **g, h**, The p44-XPD (**g**), eIF4G-eIF4A (PDB ID 2V5X ¹⁴, **h**), and Gle1-Dbp5 (PDB ID 3RRN ¹⁵, **h**) interactions occur on the same regions of RecA2 of the respective helicases. eIF4G and Gle1 additionally interact with RecA1 of eIF4A and Dbp5, respectively. **i, j**, Comparison of human XPB with the archaeal ssoRad54 SWI2/SNF2 ATPase (grey; bound double stranded DNA in red; PDB ID 1Z63) ²¹. RecA2 of the SWI2/SNF2 ATPase was superposed individually on XPB due to conformational differences. XPB shows clear homology to SWI2/SNF2 ATPases, including a domain insertion in ssoRad54 (light brown) that occurs in the same place as the XPB thumb motif (brown) (**i**) and the presence of a Gln/Asn-Trp motif in helicase motif Ia ¹⁶ (**j**). **k**, XPB RecA2 is colored using a spectrum from blue (N-terminus) to red (C-terminus). Black arrows (outwards movement and rotation) indicate the domain motions required to achieve the arrangement observed for the RecA-like domains in an archaeal SWI2/SNF2 ATPase bound to DNA (PDB ID 1Z63) ²¹. This motion would lead to extensive breakage of protein interfaces in TFIIH (regions losing contact are indicated by red arrows in right panel).

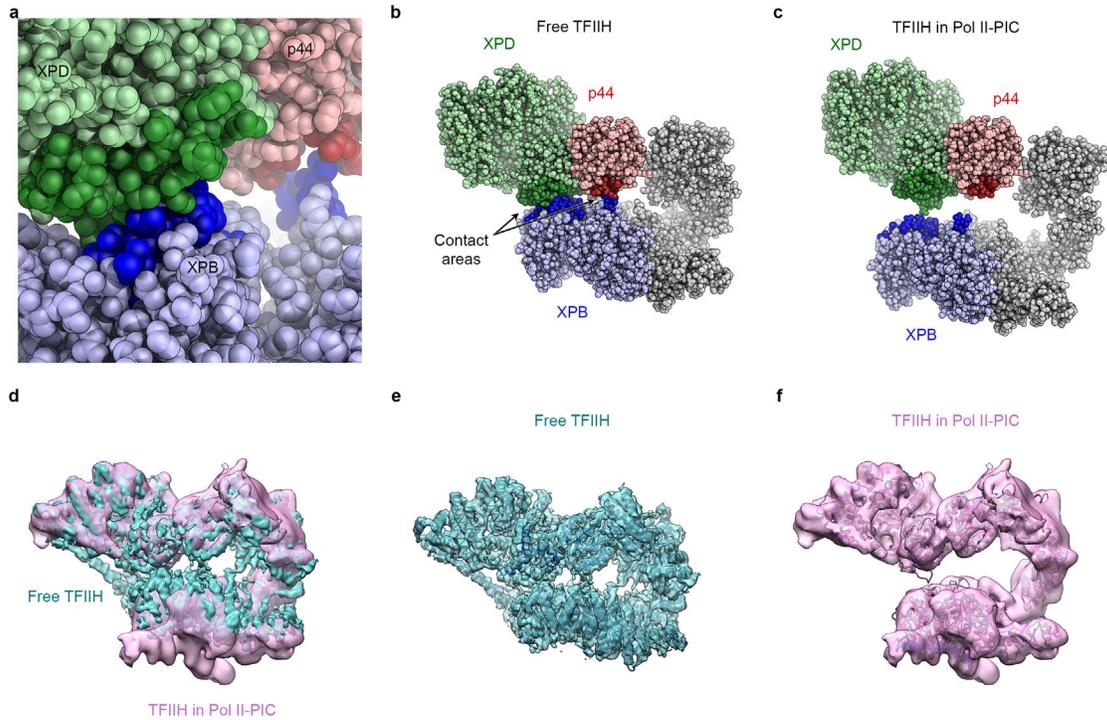


Extended Data Figure 7 | Modeling of the DRD-like domain in human XPB. a-i, Views of a fragment of PDB ID 2BOP found by an unbiased molecular replacement search of the density eventually assigned to the XPB DRD-like domain (**a-c**, see Methods), the archaeal XPB DRD (PDB ID 2FWR ²⁰; **d-f**), and the HARP1 domain (PDB ID 4O66 ⁴⁹; **g-i**) fitted into the TFIIH cryo-EM map. The fits reveal an excellent agreement of the β -sheet and the connecting α -helices with the density. **j-l**, The molecular model of TFIIH is shown in the same orientations as the above panels to show the structure of the human XPB DRD-like domain in the cryo-EM map. **m**, Comparison of the overall orientation of the DRD in archaeal XPB (cyan) and the DRD-like domain in human XPB (blue) with respect to their RecA-like domains. **n**, It has not been conclusively determined if the DRD-like domain in human XPB binds DNA. A hypothetical model obtained by superposition of the MutS DRD bound to DNA (PDB ID 1EWQ) ⁶⁵ onto the XPB DRD-like domain suggests that DRD-bound DNA would approach XPD in TFIIH. MAT1 overlaps with the DNA bound in this orientation, indicating that if the XPB DRD-like domain binds DNA as suggested by this model, it may occur in a context where MAT1 is not present in the complex, such as during NER.



Extended Data Figure 8 | Analysis of MAT1 and the remainder of the CAK

subcomplex. a, Two negative stain class averages showing clear density for the CAK subcomplex (top row) and schematic representation of the position of the CAK, MAT1, XPB, and XPD in these class averages (bottom row). **b**, Schematic representation of the position and range of motion of the CAK subcomplex inferred from 2D negative stain class averages like those shown in **a**. This high degree of structural flexibility of the CAK subcomplex precludes its visualization in the high-resolution cryo-EM map of TFIID. **c**, Negative stain 3D reconstructions of TFIID complexes including density for the CAK and data processing strategy used to obtain these reconstructions. All processing was done in RELION³⁸. The coordinate model of TFIID is fitted into three of the obtained classes. The long helix and helical bundle (Fig. 3a, b) localize to the region where the CAK and core subcomplexes interact. **d**, Unassigned density near the three-helix bundle of MAT1 is colored in blue. The map shown was obtained by signal subtracted classification⁴³ for this region of XPD/MAT1 (Extended Data Fig. 2c) and subsequent refinement. **e**, Size comparison between the N-terminal MAT1 ring finger domain⁶⁶ (PDB ID 1G25) and the unassigned density near MAT1 and the XPD ARCH domain. **f, g**, Comparison of the density in the Pol II-PIC⁵ attributed to the CAK (light yellow) to the density observed in the negative stain reconstructions of free TFIID (pink). The CAK density in free TFIID is close to the corresponding density in the context of the Pol II-PIC.



Extended Data Figure 9 | Conformational rearrangements of TFIIH. **a**, Contact areas between XPB (blue) and XPD (green) or p44 (red) are shown in space filling representation to highlight the extent of the interactions. The view is otherwise identical to the close-up in Fig. 4a. Unassigned secondary structure elements (chain Z in the atomic model) are not shown for clarity. **b**, Coordinate model of free TFIIH shown in sphere representation. Colors as in **a**. MAT1 and unassigned secondary structure elements in chain Z are not shown for clarity and ease of comparison. **c**, Same as **b**, but atomic coordinates re-fitted into the cryo-EM map of TFIIH in the context of the Pol II-PIC⁵ (EMD-8131). No contacts from XPB to XPD and p44 are observed in this complex. **d**, The comparison of the cryo-EM maps of free TFIIH (teal) and TFIIH in the context of the Pol II-PIC⁵ (EMD-8131) shows that the dimensions of the two reconstructions are almost identical along the long axis of TFIIH (horizontal), but clearly differ along the short axis (vertical) due to TFIIH opening within the DNA-engaged PIC. **e**, Model of free TFIIH in its cryo-EM map. **f** Model of TFIIH in the Pol II-PIC map⁵ (PDB ID 5IVW, EMD-8131).

Extended Data Table 1 | Summary of components included in the TFIID molecular model.

Protein	Chain ID	Size* (aa)	Modeled residues	Sequence accession code	Alternative names	Comments
XPB	A	782	221-256, 266-728	P19447	ERCC3	Residues 221-256 sequence unassigned; deposited as UNK [†] . Structural elements tentatively assigned to the XPB NTE are contained in chain X.
XPD	B	760	11-734	P18074	ERCC2	4Fe4S-cluster: SF4, residue 1000.
p62	-	548	-	P32780	GTF2H1	No density unambiguously assigned.
p52	D	462	378-458	Q92759	GTF2H4	Structural elements tentatively assigned to the p52 N-terminal helical domain are contained in chain Y.
p44	E	395	57-240	Q13888	GTF2H2	
p34	F	308	6-72, 99-122, 129-242	Q13889	GTF2H3	
p8	G	71	2-67	Q6ZYL4	GTF2H5, TTDA	
CDK7	-	346	-	P50613	MO15	Not visualized.
Cyclin H	-	323	-	P51946	CCNH	Not visualized.
MAT1	J	309	45-168	P51948	MNAT1	Sequence register unassigned; Deposited as UNK [†] .
-	X	-	7 unassigned fragments	-	-	Unassigned secondary structure elements in XPB NTE region; deposited as UNK [†] .
-	Y	-	19 unassigned fragments	-	-	Unassigned secondary structure elements in p52 helical domain region; deposited as UNK [†] .
-	Z	-	16 unassigned fragments	-	-	Unassigned secondary structure elements; deposited as UNK [†] .

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

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