

# Structures of mammalian RNA polymerase II pre-initiation complexes

<https://doi.org/10.1038/s41586-021-03554-8>

Shintaro Aibara<sup>1,2</sup>, Sandra Schilbach<sup>1,2</sup> & Patrick Cramer<sup>1✉</sup>

Received: 26 January 2021

Accepted: 14 April 2021

Published online: 26 April 2021

 Check for updates

The initiation of transcription is a focal point for the regulation of gene activity during mammalian cell differentiation and development. To initiate transcription, RNA polymerase II (Pol II) assembles with general transcription factors into a pre-initiation complex (PIC) that opens promoter DNA. Previous work provided the molecular architecture of the yeast<sup>1–9</sup> and human<sup>10,11</sup> PIC and a topological model for DNA opening by the general transcription factor TFIID<sup>12–14</sup>. Here we report the high-resolution cryo-electron microscopy structure of PIC comprising human general factors and *Sus scrofa domestica* Pol II, which is 99.9% identical to human Pol II. We determine the structures of PIC with closed and opened promoter DNA at 2.5–2.8 Å resolution, and resolve the structure of TFIID at 2.9–4.0 Å resolution. We capture the TFIID translocase XPB in the pre- and post-translocation states, and show that XPB induces and propagates a DNA twist to initiate the opening of DNA approximately 30 base pairs downstream of the TATA box. We also provide evidence that DNA opening occurs in two steps and leads to the detachment of TFIID from the core PIC, which may stop DNA twisting and enable RNA chain initiation.

Structural studies of Pol II initiation have focused on the yeast and human PIC<sup>12,13,15,16</sup>. Cryo-electron microscopy (cryo-EM) structures of the human PIC were obtained at medium resolution, and a model of the core PIC (cPIC) was derived that contained Pol II, the TATA box-binding protein (TBP), and the general transcription factors TFIIA, TFIIB, TFIIE and TFIIIF<sup>10</sup>. The human cPIC resembled the yeast cPIC structure<sup>3,4</sup> that was obtained by cryo-EM and protein–protein cross-linking analysis<sup>17</sup>, and confirmed an earlier model based on the crystal structure of Pol II–TFIIB<sup>1</sup>. TFIID was also localized in the human PIC<sup>10</sup>, which supported the model that the ATP-dependent translocase subunit XPB and its yeast counterpart Ssl2 unwind DNA from a distance by applying torque to downstream DNA<sup>6,14,18,19</sup>. The work was later extended to the human cPIC structure at a resolution of 3.9 Å<sup>11</sup>. However, portions of the cPIC remained unresolved and TFIID remained at low resolution. The structure of human TFIID was previously solved at high resolution only in free form<sup>20,21</sup> or in a functionally unrelated complex with a DNA repair factor<sup>22</sup>.

Here we report the long-sought high-resolution structure of the PIC containing mammalian Pol II and human general transcription factors including TFIID. We resolve the closed and natively opened promoter states of the PIC and define the structural mechanism that TFIID uses to open DNA. We also provide structural evidence that DNA opening starts approximately 30 base pairs (bp) downstream of the TATA box, and describe a mechanism that explains why RNA chain initiation begins at a fixed distance from the TATA box, rather than further downstream, as observed in the yeast system.

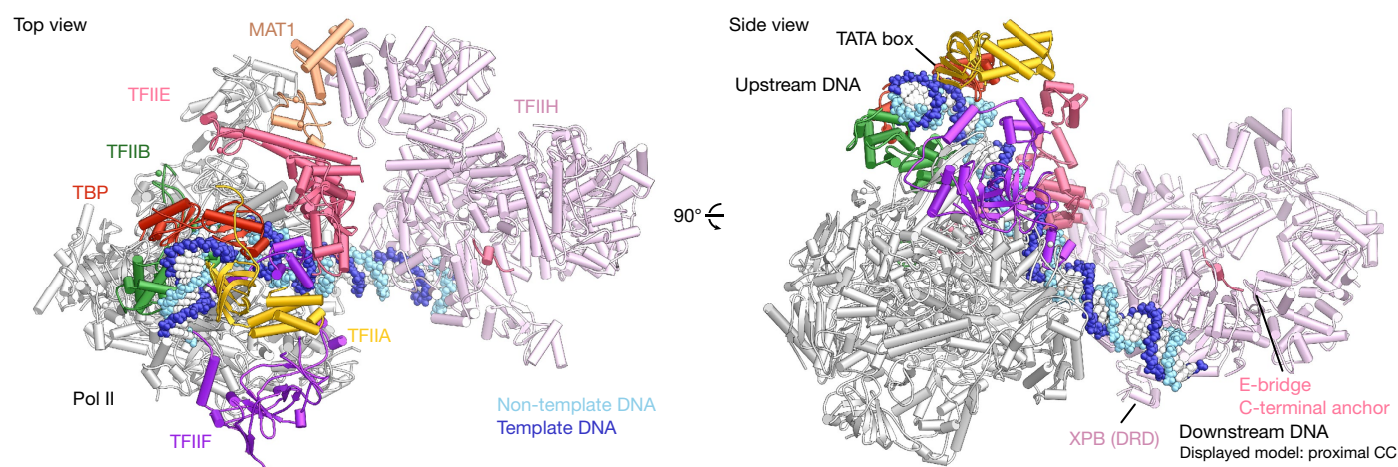
## High-resolution PIC structures

We prepared recombinant versions of human general transcription factors including TFIID<sup>22</sup> and purified endogenous Pol II from *Sus scrofa*

*domestica* (Extended Data Fig. 1a, e, Supplementary Fig. 1, Methods). We then assembled the PIC on the adenovirus major late promoter (AdMLP) and collected a single-particle cryo-EM dataset (Extended Data Fig. 1b, Supplementary Table 1). Particles that contain the complete PIC were identified by three-dimensional (3D) classification and reconstructions were improved by focused refinement and assisted model building (Extended Data Figs. 1c, 2, 3, Methods). This led to two structures of the closed promoter complex (CC) where the cPIC portion was resolved at resolutions of 2.8 Å and 2.9 Å that deviated slightly in the location of the upstream complex that contains TBP, TFIIA, TFIIB and DNA (Extended Data Fig. 4). The upstream complex was observed either in a proximal location on the Pol II surface, or in a more distal position, shifted by up to 7 Å away from Pol II. TFIID was resolved at resolutions of 3.3 Å and 4.1 Å for its XPB-containing part (XPB, p52 and p8) and its XPD-containing part (XPD, p62, p44, p34 and MAT1), respectively.

We also collected a PIC dataset in the presence of ADP·BeF<sub>3</sub>, which mimics the transition state of ATP hydrolysis and was thought to reduce flexibility of TFIID (Extended Data Fig. 1b, Supplementary Table 1). In contrast to the first dataset, which mainly provided CC classes, we now also obtained reconstructions for the open promoter complex (OC). This showed that the PIC had actively opened the fully matched, closed DNA duplex. Focused refinement of particles classified by DNA conformation resulted in reconstructions of the cPIC, the XPB-containing part of TFIID, and the XPD-containing part of TFIID at resolutions of 2.5 Å, 2.9 Å, and 4.0 Å, respectively (Extended Data Fig. 5). This led to a refined atomic model of the OC on the basis of a single reconstruction of the entire PIC (Extended Data Fig. 6). Compared with the most complete previous human PIC model<sup>11</sup>, we could include approximately 500 additional amino acid residues (Methods).

<sup>1</sup>Department of Molecular Biology, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany. <sup>2</sup>These authors contributed equally: Shintaro Aibara, Sandra Schilbach. ✉e-mail: patrick.cramer@mpibpc.mpg.de



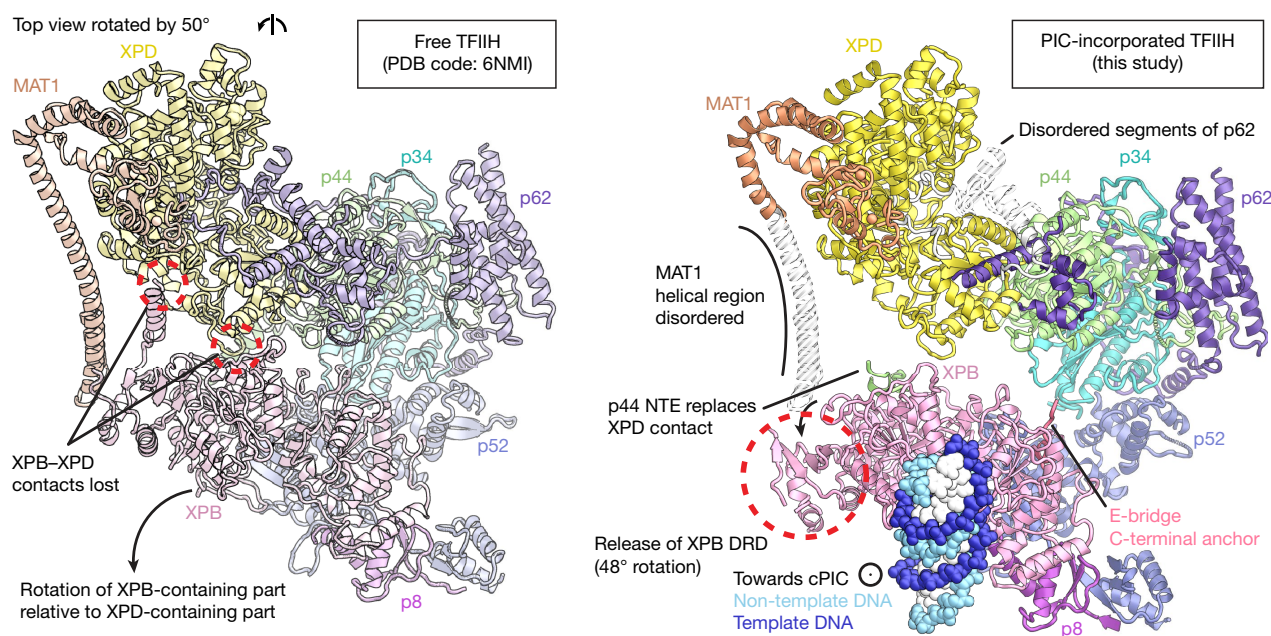
**Fig. 1 | Structure of the mammalian PIC.** Structure of the CC viewed from the top and side. The same colour code is used for proteins and nucleic acids throughout the manuscript. DRD, damage recognition domain.

## TFIIH structure and interactions

Our structure resolves all chains of TFIIH, except for the flexibly linked and mobile CDK7–cyclin H pair (Fig. 1, Supplementary Video 1). The observed TFIIH conformation differs from the structure of free TFIIH<sup>21</sup> (Fig. 2). In particular, the XPD–XPB interface is broken<sup>21</sup>, as observed in the yeast PIC<sup>6</sup>. The XPD-containing part of TFIIH rotates by approximately 10° around an axis running through subunit p52. This rotation dislodges XPD residues 499–501 from the XPB damage recognition domain and generates space that is partly occupied by the N-terminal extension of p44. The resulting contact between p44 and XPB probably stabilizes TFIIH in its transcription-related conformation, which strongly deviates from the DNA repair-related conformation<sup>22</sup>. The conformational changes in TFIIH are coupled to displacement of the long helical domain in MAT1, enabling an approximately 50° rotation of the damage recognition domain and interaction of XPB with DNA<sup>22</sup>. In summary, the MAT1 helical domain

has to detach to liberate XPB and allow TFIIH to bind DNA within the PIC (Supplementary Fig. 2).

Our PIC structure also defines the interfaces between the cPIC and TFIIH. The major interface is formed between the RING finger domain of MAT1 and the Pol II stalk and TFIIE (Extended Data Fig. 7a). This interface is consistent with a biochemically defined role of TFIIE in anchoring the TFIIH kinase module on the PIC<sup>23</sup>. The RING finger domain of MAT1 binds the oligonucleotide-binding fold of the stalk subunit RPB7 and the E-linker of TFIIE (Extended Data Fig. 7a). This contrasts with the yeast PIC, in which the RING finger domain of the MAT1 counterpart Tfb3 contacts both stalk subunits, Rpb7 and Rpb4. The TFIIE element E-dock that was previously described in the yeast PIC<sup>6</sup> is also visible in the human PIC density, but could not be assigned with confidence (Extended Data Fig. 7a). A conserved<sup>6</sup> interface is also formed between XPB and the E-bridge element of the large TFIIE subunit TFIIE $\alpha$  (Extended Data Fig. 7b). In summary, the contacts between the cPIC and TFIIH defined in the yeast PIC



**Fig. 2 | Changes in TFIIH upon PIC formation.** Comparison between TFIIH in free form<sup>21</sup> (PDB code 6NMI) and TFIIH within the PIC (this study). Subunits of TFIIH are coloured individually and labelled. The MAT1 helical region becomes

disordered upon PIC incorporation (dashed outline). NTE, N-terminal extension.

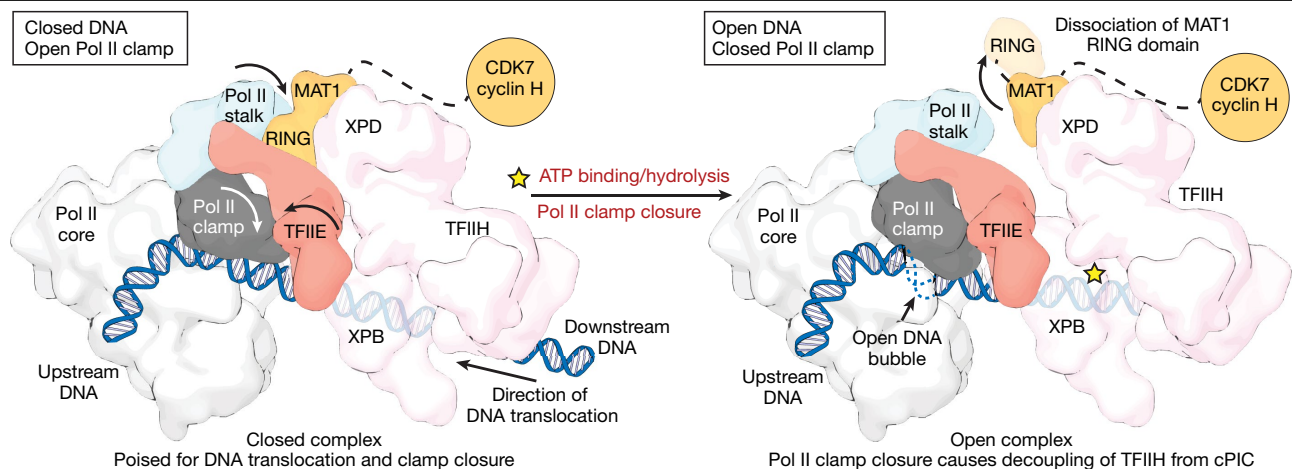




are largely conserved in the human PIC, but some minor differences also occur.

The addition of ADP-BeF<sub>3</sub> to the PIC sample changed the distribution of particles with closed or opened DNA (Extended Data Fig. 5). In the CC and OC states derived from the two distinct preparations, the TFIID translocase XPB adopts different conformations (Fig. 3a). In the CC, the ATPase domain adopts an open, pre-translocated state with an empty

In contrast to the human PIC, the yeast PIC can open a subset of promoters without the help of TFIIF<sup>3,5,6</sup>. This functional difference between human and yeast PICs may be related to differences in the dynamics of the Pol II clamp and in the structure of TFIIB. Whereas the position of the yeast Pol II clamp remains unchanged during the CC-to-OC transition<sup>3,6,25</sup>, the clamp of mammalian Pol II rotates to close over DNA<sup>11,16</sup>. Consistent with this, we observe that the clamp rotates by around 15°



**Fig. 4 | TFIID detachment during OC formation.** Cartoon schematic showing movements of the clamp, stalk and TFIIE during the CC-to-OC transition. As clamp closure and DNA opening occurs, the MAT1 RING finger domain is displaced, thus detaching TFIID from the cPIC.

when the CC is converted to the natively opened OC (Extended Data Fig. 9b, Supplementary Video 3). Superposition of our structures of CC and OC shows that an open clamp is required to accommodate closed DNA in the human PIC (Extended Data Fig. 9b). Clamp closure leads to charge-based contacts between the clamp and DNA that stabilize the OC and facilitate the second step of DNA opening. Indeed, we observe open DNA only in PIC particles with a closed clamp, and clamp closure is therefore a hallmark of human OC formation.

Clamp closure and formation of the OC induce ordering of the TFIIB elements B-reader and B-linker, which traverse the Pol II cleft and are mobile in the CC (Extended Data Fig. 9c). Human TFIIB differs from its yeast counterpart<sup>2</sup> in that it lacks the  $\alpha$ -helical regions of the B-reader and B-linker (Extended Data Fig. 9c). The absence of these two helices was noted<sup>11</sup>, but refined atomic models of the B-reader and B-linker could only now be obtained (Extended Data Fig. 9c). The B-linker helix is known to function in DNA opening in the related archaeal PIC<sup>1</sup>, which lacks TFIID and opens DNA without a translocase activity. Thus, the absence of the B-linker helix in human TFIIB can explain why the human PIC cannot open DNA without XPB translocase activity.

### TFIID detachment and transcription start site

Classification of our data further indicated that TFIID can be either attached or detached from the cPIC. In the OC class of particles, TFIID is detached from the cPIC even when we deliberately tried to enrich for particles with attached TFIID (Fig. 4, Extended Data Figs. 2, 5). In this reconstruction, the contact of the MAT1 RING finger domain with the Pol II stalk and TFIIE is broken, the RING finger domain is mobile, and TFIID associates with the cPIC only indirectly through DNA and flexible tethers. Particles from all other classes did not exhibit such TFIID detachment and the RING finger domain was always detected, whether ADP-BeF<sub>3</sub> was present or absent (Extended Data Fig. 9d). Much fewer particles with detached TFIID could be identified in CC classes. Therefore, the CC-to-OC transition is associated with TFIID detachment from the cPIC. The detachment of the MAT1 RING finger domain may be stimulated by the closure of the Pol II clamp (Extended Data Fig. 9b, Supplementary Video 3).

TFIID detachment from the cPIC is predicted to stop further DNA twisting because the XPB translocase motor uses the TFIID–cPIC contact as a stator and cannot generate torque on DNA without a stator. Continued DNA twisting is, however, required for scanning of downstream DNA for the transcription start site, which is a feature of the yeast PIC<sup>32</sup>. The observed TFIID detachment in the OC state may therefore explain why the mammalian PIC, in contrast to the yeast PIC, does not

scan downstream DNA. In strong support of this model, the TFIID kinase module is known to be required for DNA scanning, and its deletion abolishes scanning and leads to a human-like use of the transcription start site in the yeast system<sup>33</sup>. TFIID may detach more easily from the cPIC in the human system because contacts between MAT1 and the Pol II stalk are less extensive than the corresponding contacts in the yeast PIC (Extended Data Fig. 7b). In summary, TFIID detachment from the cPIC occurs after formation of the OC in the human system and is predicted to disable downstream DNA scanning and trigger the initiation of RNA synthesis at around 30–35 bp downstream of the TATA box.

### Conclusion

The long-sought high-resolution structure of the mammalian PIC in different functional states provides detailed insights into the mechanism of transcription initiation. The structure of the mammalian PIC will enable the determination of structures of larger assemblies and mechanistic studies of the regulation of transcription initiation. As a first step in this direction, the accompanying paper<sup>34</sup> reports the structure of the PIC bound by the human regulatory coactivator complex Mediator.

### Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-021-03554-8>.

1. Kostrewa, D. et al. RNA polymerase II–TFIIB structure and mechanism of transcription initiation. *Nature* **462**, 323–330 (2009).
2. Sainsbury, S., Niesser, J. & Cramer, P. Structure and function of the initially transcribing RNA polymerase II–TFIIB complex. *Nature* **493**, 437–440 (2013).
3. Plaschka, C. et al. Transcription initiation complex structures elucidate DNA opening. *Nature* **533**, 353–358 (2016).
4. Plaschka, C. et al. Architecture of the RNA polymerase II–Mediator core initiation complex. *Nature* **518**, 376–380 (2015).
5. Dienemann, C., Schwalb, B., Schilbach, S. & Cramer, P. Promoter distortion and opening in the RNA polymerase II cleft. *Mol. Cell* **73**, 97–106 (2019).
6. Schilbach, S. et al. Structures of transcription pre-initiation complex with TFIID and Mediator. *Nature* **551**, 204–209 (2017).
7. Liu, X., Bushnell, D. A., Wang, D., Calero, G. & Kornberg, R. D. Structure of an RNA polymerase II–TFIIB complex and the transcription initiation mechanism. *Science* **327**, 206–209 (2010).
8. Robinson, P. J. et al. Structure of a complete mediator–RNA polymerase II pre-initiation complex. *Cell* **166**, 1411–1422 (2016).
9. Murakami, K. et al. Structure of an RNA polymerase II preinitiation complex. *Proc. Natl Acad. Sci. USA* **112**, 13543–13548 (2015).



10. He, Y., Fang, J., Taatjes, D. J. & Nogales, E. Structural visualization of key steps in human transcription initiation. *Nature* **495**, 481–486 (2013).
11. He, Y. et al. Near-atomic resolution visualization of human transcription promoter opening. *Nature* **533**, 359–365 (2016).
12. Sainsbury, S., Bernecky, C. & Cramer, P. Structural basis of transcription initiation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* **16**, 129–143 (2015).
13. Hahn, S. & Young, E. T. Transcriptional regulation in *Saccharomyces cerevisiae*: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. *Genetics* **189**, 705–736 (2011).
14. Fishburn, J., Tomko, E., Galburt, E. & Hahn, S. Double-stranded DNA translocase activity of transcription factor TFIIF and the mechanism of RNA polymerase II open complex formation. *Proc. Natl Acad. Sci. USA* **112**, 3961–3966 (2015).
15. Nogales, E., Louder, R. K. & He, Y. Structural insights into the eukaryotic transcription initiation machinery. *Annu. Rev. Biophys.* **46**, 59–83 (2017).
16. Hantsche, M. & Cramer, P. Conserved RNA polymerase II initiation complex structure. *Curr. Opin. Struct. Biol.* **47**, 17–22 (2017).
17. Mühlbacher, W. et al. Conserved architecture of the core RNA polymerase II initiation complex. *Nat. Commun.* **5**, 4310 (2014).
18. 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).
19. Kim, T. K., Ebright, R. H. & Reinberg, D. Mechanism of ATP-dependent promoter melting by transcription factor IIH. *Science* **288**, 1418–1422 (2000).
20. Greber, B. J. et al. The cryo-electron microscopy structure of human transcription factor IIH. *Nature* **549**, 414–417 (2017).
21. Greber, B. J., Toso, D. B., Fang, J. & Nogales, E. The complete structure of the human TFIIF core complex. *eLife* **8**, e44771 (2019).
22. Kocic, G. et al. Structural basis of TFIIF activation for nucleotide excision repair. *Nat. Commun.* **10**, 2885 (2019).
23. Compe, E., Genes, C. M., Braun, C., Coin, F. & Egly, J. M. TFIIIE orchestrates the recruitment of the TFIIF kinase module at promoter before release during transcription. *Nat. Commun.* **10**, 2084 (2019).
24. Farnung, L., Ochmann, M. & Cramer, P. Nucleosome-CHD4 chromatin remodeler structure maps human disease mutations. *eLife* **9**, e56178 (2020).
25. Li, M. et al. Mechanism of DNA translocation underlying chromatin remodelling by Snf2. *Nature* **567**, 409–413 (2019).
26. Wang, W., Carey, M. & Gralla, J. D. Polymerase II promoter activation: closed complex formation and ATP-driven start site opening. *Science* **255**, 450–453 (1992).
27. Holstege, F. C., van der Vliet, P. C. & Timmers, H. T. Opening of an RNA polymerase II promoter occurs in two distinct steps and requires the basal transcription factors IIE and IIH. *EMBO J.* **15**, 1666–1677 (1996).
28. Giardina, C. & Lis, J. T. DNA melting on yeast RNA polymerase II promoters. *Science* **261**, 759–762 (1993).
29. Tomko, E. J., Fishburn, J., Hahn, S. & Galburt, E. A. TFIIF generates a six-base-pair open complex during RNAP II transcription initiation and start-site scanning. *Nat. Struct. Mol. Biol.* **24**, 1139–1145 (2017).
30. Ren, D., Lei, L. & Burton, Z. F. A region within the RAP74 subunit of human transcription factor IIF is critical for initiation but dispensable for complex assembly. *Mol. Cell. Biol.* **19**, 7377–7387 (1999).
31. Tomko, E. J. et al. The role of XPB/XPD dsDNA translocase processivity in transcription start-site scanning. *J. Mol. Biol.* <https://doi.org/10.1016/j.jmb.2021.166813> (2021).
32. Qiu, C. et al. Universal promoter scanning by Pol II during transcription initiation in *Saccharomyces cerevisiae*. *Genome Biol.* **21**, 132 (2020).
33. Murakami, K. et al. Uncoupling promoter opening from start-site scanning. *Mol. Cell* **59**, 133–138 (2015).
34. Rengachari, S., Schilbach, S., Aibara, S., Dienemann, C. & Cramer, P. Structure of the human Mediator–RNA polymerase II pre-initiation complex. *Nature* <https://doi.org/10.1038/s41586-021-3555-7> (2021).

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

© The Author(s), under exclusive licence to Springer Nature Limited 2021

## Methods

### Data reporting

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

### Cloning and protein expression

DNA segments encoding full-length *Homo sapiens* TFIIA- $\alpha$ , TFIIA- $\beta$ , TFIIB, TBP, TFIIE- $\alpha$ , TFIIE- $\beta$ , TFIIF- $\alpha$  (Rap74) and TFIIF- $\beta$  (Rap30) were purchased as gBlocks gene fragments from Integrated DNA Technologies (IDT). Sequences for TFIIA- $\alpha$ , TFIIA- $\beta$ , TFIIE- $\alpha$ , TFIIE- $\beta$ , Rap74 and Rap30 were obtained as codon-optimized variants, sequences for TFIIA- $\alpha$  and TFIIE- $\beta$  were directly fused to C-terminal 6 $\times$ His-tags and the sequence for TBP was fused to a N-terminal 6 $\times$ His-tag followed by a cleavage site for human rhinovirus 3C (HRV-3C) protease, respectively. DNA fragments were amplified by PCR and transferred into various types of *Escherichia coli* or insect cell expression vectors by ligation-independent cloning (LIC), In-Fusion cloning (Takara Bio) or cloning by restriction-enzyme digest. The final constructs comprised TFIIA- $\alpha$ -6xHis and TFIIA- $\beta$  in a 438-A vector<sup>35</sup> (Addgene 55218), 6xHis-3C-TFIIB in a pOPINF vector<sup>36</sup> (Addgene 26042), 6xHis-3C-TBP in a 438-A vector, TFIIE- $\alpha$  and TFIIE- $\beta$ -6xHis in a pETDuet-1 vector (Novagen) and 10xHis-8xArg-SUMO-3C-Rap74 and Rap30 in a modified pETDuet-1 vector (pAHS3C)<sup>3</sup>. Plasmid sequences are available upon request.

TFIIB was expressed in LOBSTR-BL21(DE3)-RIL *E. coli* cells (Kerafast) by auto-induction essentially as previously reported<sup>37</sup>. The  $\alpha$ - and  $\beta$ -subunits of TFIIE and TFIIF were co-expressed in BL21-Codon Plus(DE3)-RIL *E. coli* cells (Agilent) for 20 h at 18 °C, respectively. Cells were collected by centrifugation (7,900g, 10 min, 4 °C) and resuspended in lysis buffer (500 mM KCl, 25 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid/KOH (HEPES/KOH) pH 7.5, 15% glycerol (v/v), 30 mM imidazole, 5 mM  $\beta$ -mercaptoethanol, 0.284  $\mu$ g ml<sup>-1</sup> leupeptin, 1.37  $\mu$ g ml<sup>-1</sup> pepstatin A, 0.17 mg ml<sup>-1</sup> phenylmethylsulfonyl fluoride (PMSF), 0.33 mg ml<sup>-1</sup> benzamidine). The resulting cell suspensions were flash-cooled in liquid nitrogen and stored at -80 °C until purification.

Preparation of bacmids containing the gene sequences for TFIIA- $\alpha$  and - $\beta$  subunits or for TBP, production of the respective insect cell viruses of the V0 and V1 stage in Sf9 insect cells and protein expression in Hi5 insect cells were performed as described<sup>6</sup>. Cells were collected by centrifugation (238g, 45 min, 4 °C) and resuspended in lysis buffer (500 mM KCl, 25 mM HEPES/KOH pH 7.5, 10% glycerol (v/v), 30 mM imidazole, 5 mM  $\beta$ -mercaptoethanol, 0.284  $\mu$ g ml<sup>-1</sup> leupeptin, 1.37  $\mu$ g ml<sup>-1</sup> pepstatin A, 0.17 mg ml<sup>-1</sup> PMSF, 0.33 mg ml<sup>-1</sup> benzamidine). The resulting cell suspensions were flash-cooled in liquid nitrogen and stored at -80 °C until purification.

### Protein purification

All purification procedures were performed at 4 °C unless stated otherwise. Purification buffers were filtered and degassed before use. Peptide identity of the final products was confirmed by mass spectrometry. Pol II was obtained from *S. scrofa domesticus* thymus material as reported using resin-conjugated 8WG16 antibodies<sup>38,39</sup>. TFIIF core and kinase modules were purified as previously described<sup>22</sup> with minor alterations.

TFIIA, TFIIB, TBP, TFIIE and TFIIF were prepared by consecutive steps of affinity chromatography, ion-exchange chromatography and size-exclusion chromatography. Generally, frozen cell pellets were thawed in a water bath at 25 °C. *E. coli* cells were lysed by sonication, insect cells were supplemented with catalytic amounts of DNase I and lysed with an EmulsiFlex-C5 cell disruptor (Avestin) (3 passages, 12,000 psi). Cell lysates were cleared by centrifugation (79,000g; 60 min) and the soluble fraction was filtered through 0.45- $\mu$ m syringe filters (Merck Millipore) before affinity chromatography.

TFIIA-containing lysate was applied to a GE HisTrap HP (5 ml) column (GE Healthcare) that had been pre-equilibrated in buffer H<sub>A</sub>-30 (500 mM KCl, 25 mM HEPES/KOH pH 7.5, 10% glycerol (v/v), 30 mM imidazole, 5 mM  $\beta$ -mercaptoethanol, 0.284  $\mu$ g ml<sup>-1</sup> leupeptin, 1.37  $\mu$ g ml<sup>-1</sup> pepstatin A, 0.17 mg ml<sup>-1</sup> PMSF, 0.33 mg ml<sup>-1</sup> benzamidine). The column was washed with 12 column volumes (CV) of buffer H<sub>A</sub>-30 and 5 CV of buffer H<sub>A</sub>-HS (1 M KCl, 25 mM HEPES/KOH pH 7.5, 10% glycerol (v/v), 30 mM imidazole, 5 mM  $\beta$ -mercaptoethanol, 0.284  $\mu$ g ml<sup>-1</sup> leupeptin, 1.37  $\mu$ g ml<sup>-1</sup> pepstatin A, 0.17 mg ml<sup>-1</sup> PMSF, 0.33 mg ml<sup>-1</sup> benzamidine), followed by another 5 CV of buffer H<sub>A</sub>-30 and 6 CV of buffer HEP<sub>A</sub>-100 (100 mM KCl, 25 mM HEPES/KOH pH 7.5, 5% glycerol (v/v), 5 mM  $\beta$ -mercaptoethanol). Bound protein was eluted with a linear gradient of 0–100% buffer H<sub>A</sub>-500 (100 mM KCl, 25 mM HEPES/KOH pH 7.5, 5% glycerol (v/v), 500 mM imidazole, 5 mM  $\beta$ -mercaptoethanol) in 12 CV. Peak fractions were pooled and applied to a GE HiTrap Heparin HP (5 ml) column (GE Healthcare) pre-equilibrated in buffer HEP<sub>A</sub>-100. Fractions of the flow through and a subsequent wash with 15 CV of buffer HEP<sub>A</sub>-100 were collected, pooled and applied to a GE HiTrap Q HP (1 ml) column (GE Healthcare) pre-equilibrated in buffer HEP<sub>A</sub>-100. The column was washed with 10 CV of buffer HEP<sub>A</sub>-100 and bound protein was eluted with a step to 15% buffer HQ<sub>A</sub>-1000 (1 M KCl, 25 mM HEPES/KOH pH 7.5, 5% glycerol (v/v), 5 mM  $\beta$ -mercaptoethanol), followed by a linear gradient of 15–75% buffer HQ<sub>A</sub>-1000 in 50 CV. Peak fractions were pooled, concentrated with an Amicon Millipore 15 ml 30-kDa MWCO centrifugal device (Amicon) and applied to a GE Superdex200 10/300 GL size exclusion column (GE Healthcare) pre-equilibrated in size exclusion buffer (200 mM KCl, 25 mM HEPES/KOH, pH 7.5, 5% glycerol (v/v), 2 mM tris(2-carboxyethyl)phosphine (TCEP)). Peak fractions containing a stoichiometric dimer of TFIIA- $\alpha$  and TFIIA- $\beta$  were pooled, concentrated to 10 mg ml<sup>-1</sup> using an Amicon Millipore 6 ml 30-kDa MWCO centrifugal device (Amicon), aliquots were flash-cooled in liquid nitrogen and stored at -80 °C. Typical yields were in the range of 5 mg per litre of Hi5 insect cell culture.

TFIIB-containing lysate was applied to a GE HisTrap HP (5 ml) column (GE Healthcare) that had been pre-equilibrated in buffer H<sub>B</sub>-30 (500 mM KCl, 25 mM HEPES/KOH pH 7.5, 15% glycerol (v/v), 30 mM imidazole, 5  $\mu$ M ZnCl<sub>2</sub>, 5 mM  $\beta$ -mercaptoethanol, 0.284  $\mu$ g ml<sup>-1</sup> leupeptin, 1.37  $\mu$ g ml<sup>-1</sup> pepstatin A, 0.17 mg ml<sup>-1</sup> PMSF, 0.33 mg ml<sup>-1</sup> benzamidine). The column was washed with 12 CV of buffer H<sub>B</sub>-30, 5 CV of buffer H<sub>B</sub>-HS (1 M KCl, 25 mM HEPES/KOH pH 7.5, 15% glycerol (v/v), 30 mM imidazole, 5  $\mu$ M ZnCl<sub>2</sub>, 5 mM  $\beta$ -mercaptoethanol, 0.284  $\mu$ g ml<sup>-1</sup> leupeptin, 1.37  $\mu$ g ml<sup>-1</sup> pepstatin A, 0.17 mg ml<sup>-1</sup> PMSF, 0.33 mg ml<sup>-1</sup> benzamidine) and another 5 CV of buffer H<sub>B</sub>-30. Bound protein was eluted with a linear gradient of 0–100% buffer H<sub>B</sub>-500 (500 mM KCl, 25 mM HEPES/KOH pH 7.5, 15% glycerol (v/v), 500 mM imidazole, 5  $\mu$ M ZnCl<sub>2</sub>, 5 mM  $\beta$ -mercaptoethanol) in 12 CV. Peak fractions were pooled, supplemented with 0.5 mg 6 $\times$ His-HRV-3C protease and dialysed against buffer D<sub>B</sub> (200 mM KCl, 25 mM HEPES/KOH pH 7.5, 10% glycerol (v/v), 5  $\mu$ M ZnCl<sub>2</sub>, 5 mM  $\beta$ -mercaptoethanol) at 4 °C for 8 h. The cleaved protein sample was passed through a 0.2- $\mu$ m filter (Merck Millipore) and subjected to cation exchange chromatography using a GE HiTrap SP HP (1 ml) column (GE Healthcare) pre-equilibrated in buffer HSP<sub>B</sub>-200 (200 mM KCl, 25 mM HEPES/KOH pH 7.5, 10% glycerol (v/v), 5  $\mu$ M ZnCl<sub>2</sub>, 5 mM  $\beta$ -mercaptoethanol). After sample application, the column was washed with 10 CV of buffer HSP<sub>B</sub>-200 and protein was eluted with a linear gradient of 0–75% buffer HSP<sub>B</sub>-1000 (1 M KCl, 25 mM HEPES/KOH pH 7.5, 10% glycerol (v/v), 5  $\mu$ M ZnCl<sub>2</sub>, 5 mM  $\beta$ -mercaptoethanol) in 100 CV. Peak fractions were pooled, concentrated with an Amicon Millipore 6 ml 10-kDa MWCO centrifugal device (Amicon) and applied to a GE Superdex75 10/300 GL size exclusion column (GE Healthcare) pre-equilibrated in size exclusion buffer. Peak fractions containing pure TFIIB were pooled, concentrated to 4–5 mg ml<sup>-1</sup> using a Vivaspinn 500 MWCO 10-kDa centrifugal device (GE Healthcare), aliquots were flash-cooled in liquid nitrogen and stored at -80 °C. Typical yields were in the range of 0.2–0.4 mg per litre of *E. coli* cell culture.



TBP-containing lysate was applied to a GE HisTrap HP (5 ml) column (GE Healthcare) that had been pre-equilibrated in buffer H<sub>T</sub>-30 (500 mM KCl, 25 mM HEPES/KOH pH 7.5, 10% glycerol (v/v), 30 mM imidazole, 5 mM β-mercaptoethanol, 0.284 μg ml<sup>-1</sup> leupeptin, 1.37 μg ml<sup>-1</sup> pepstatin A, 0.17 mg ml<sup>-1</sup> PMSF, 0.33 mg ml<sup>-1</sup> benzamidine). The column was washed with 12 CV of buffer H<sub>T</sub>-30 and bound protein was eluted with a linear gradient of 0–100% buffer H<sub>T</sub>-500 (500 mM KCl, 25 mM HEPES/KOH pH 7.5, 10% glycerol (v/v), 500 mM imidazole, 5 mM β-mercaptoethanol) in 12 CV. TBP-containing fractions were pooled, supplemented with 0.5 mg 6xHis-HRV-3C protease and dialysed against buffer D<sub>T</sub> (150 mM KCl, 25 mM HEPES/KOH pH 7.5, 5% glycerol (v/v), 5 mM β-mercaptoethanol) at 4 °C for 8 h. The cleaved protein sample was passed through a 0.2-μm filter (Merck Millipore) and subjected to cation exchange chromatography using a GE HiTrap SP HP (1 mL) column (GE Healthcare) pre-equilibrated in buffer HSP<sub>T</sub>-150 (150 mM KCl, 25 mM HEPES/KOH pH 7.5, 5% glycerol (v/v), 5 mM β-mercaptoethanol). After sample application the column was washed with 10 CV of buffer HSP<sub>T</sub>-150 and the protein was eluted with a linear gradient of 0–50% buffer HSP<sub>T</sub>-1000 (1 M KCl, 25 mM HEPES/KOH pH 7.5, 5% glycerol (v/v), 5 mM β-mercaptoethanol) in 100 CV. Peak fractions were pooled, concentrated with an Amicon Millipore 6 mL 10 000 MWCO centrifugal device (Amicon) and applied to a GE Superose12 10/300 GL size exclusion column (GE Healthcare) pre-equilibrated in size exclusion buffer. Peak fractions containing pure TBP were pooled, concentrated to 4–5 mg ml<sup>-1</sup> using a Vivaspin 500 MWCO 10-kDa centrifugal device (GE Healthcare), aliquots were flash-cooled in liquid nitrogen and stored at –80 °C. Typical yields were in the range of 0.5 mg per litre of Hi5 insect cell culture.

*H. sapiens* TFIIIE was purified by consecutive steps of affinity chromatography, ion exchange chromatography and size exclusion chromatography. Thawed cells were lysed by sonication and the cell lysate was cleared by centrifugation (79,000g; 60 min). The soluble fraction was filtered and applied to a GE HisTrap HP (5 ml) column (GE Healthcare) pre-equilibrated in buffer E-5 (300 mM NaCl, 50 mM Tris/HCl pH 8.0, 5 mM imidazole, 2 mM 1,4-dithiothreitol, 0.284 μg ml<sup>-1</sup> leupeptin, 1.37 μg ml<sup>-1</sup> pepstatin A, 0.17 mg ml<sup>-1</sup> PMSF, 0.33 mg ml<sup>-1</sup> benzamidine). The column was washed with 10 CV of buffer E-5 and 5 CV of buffer E-15 (300 mM NaCl, 50 mM Tris-HCl pH 8.0, 15 mM imidazole, 2 mM 1,4-dithiothreitol, 0.284 μg ml<sup>-1</sup> leupeptin, 1.37 μg ml<sup>-1</sup> pepstatin A, 0.17 mg ml<sup>-1</sup> PMSF, 0.33 mg ml<sup>-1</sup> benzamidine). Protein was eluted with 5 CV buffer E-250 (300 mM NaCl, 50 mM Tris-HCl pH 8.0, 250 mM imidazole, 2 mM 1,4-dithiothreitol, 0.284 μg ml<sup>-1</sup> leupeptin, 1.37 μg ml<sup>-1</sup> pepstatin A, 0.17 mg ml<sup>-1</sup> PMSF, 0.33 mg ml<sup>-1</sup> benzamidine), pooled and dialysed against buffer D (200 mM NaCl, 50 mM Tris-HCl pH 8.0, 2% glycerol (v/v), 5 mM 1,4-dithiothreitol) for 12 h. The dialysed sample was applied to a GE HiTrap Heparin HP (5 ml) column (GE Healthcare) pre-equilibrated in buffer HEP-240 (240 mM NaCl, 50 mM Tris/HCl pH 8.0, 2% glycerol (v/v), 5 mM 1,4-dithiothreitol). The column was washed with 20 CV of buffer HEP-240 and the protein was eluted with a linear gradient of 0–70% buffer HEP-2000 (2 M NaCl, 50 mM Tris-HCl pH 8.0, 2% glycerol (v/v), 5 mM 1,4-dithiothreitol) in 15 CV. Fractions containing stoichiometric TFIIIE were pooled, concentrated with a Millipore 15 ml 10-kDa MWCO centrifugal device (Amicon) and applied to a GE Superose12 10/300 GL size exclusion column pre-equilibrated in size exclusion buffer (200 mM KCl, 20 mM HEPES/KOH pH 7.0, 5% glycerol (v/v), 3 mM 1,4-dithiothreitol). Peak fractions were pooled, concentrated to 4–5 mg ml<sup>-1</sup> using a Millipore 6 ml 10-kDa MWCO centrifugal device (Amicon), aliquoted, flash-cooled in liquid nitrogen and stored at –80 °C. Typical yields were 5 mg per litre of *E. coli* culture.

TFIIF-containing lysate was applied to a GE HisTrap HP (5 ml) column (GE Healthcare) that had been pre-equilibrated in buffer H<sub>F</sub>-50 (500 mM KCl, 25 mM HEPES/KOH pH 7.5, 10% glycerol (v/v), 50 mM imidazole, 5 mM β-mercaptoethanol, 0.284 μg ml<sup>-1</sup> leupeptin, 1.37 μg ml<sup>-1</sup> pepstatin A, 0.17 mg ml<sup>-1</sup> PMSF, 0.33 mg ml<sup>-1</sup> benzamidine). The column was washed with 12 CV of buffer H<sub>F</sub>-50, 5 CV of buffer H<sub>F</sub>-125 (1 M KCl,

25 mM HEPES/KOH pH 7.5, 10% glycerol (v/v), 125 mM imidazole, 5 mM β-mercaptoethanol, 0.284 μg ml<sup>-1</sup> leupeptin, 1.37 μg ml<sup>-1</sup> pepstatin A, 0.17 mg ml<sup>-1</sup> PMSF, 0.33 mg ml<sup>-1</sup> benzamidine) and another 5 CV of buffer H<sub>F</sub>-50. Bound protein was eluted with 5 CV of 30% buffer H<sub>F</sub>-800 (500 mM KCl, 25 mM HEPES/KOH pH 7.5, 10% glycerol (v/v), 800 mM imidazole, 5 mM β-mercaptoethanol). Peak fractions were pooled, supplemented with 0.5 mg 6xHis-HRV-3C protease and dialysed against buffer D<sub>F</sub> (100 mM KCl, 25 mM HEPES/KOH pH 7.5, 10% glycerol (v/v), 5 mM β-mercaptoethanol) at 4 °C for 8 h. The cleaved protein sample was passed through a 0.2-μm filter (Merck Millipore) and subjected to cation exchange chromatography using a GE HiTrap SP HP (1 ml) column (GE Healthcare) pre-equilibrated in buffer HSP<sub>F</sub>-100 (100 mM KCl, 25 mM HEPES/KOH pH 7.5, 10% glycerol (v/v), 5 mM β-mercaptoethanol). After sample application the column was washed with 10 CV of buffer HSP<sub>F</sub>-100 and the protein was eluted with a linear gradient of 0–20% buffer HSP<sub>F</sub>-2000 (2 M KCl, 25 mM HEPES/KOH pH 7.5, 10% glycerol (v/v), 5 mM β-mercaptoethanol) in 60 CV. Fractions containing stoichiometric TFIIIF subunits Rap74 and Rap30 were pooled, concentrated with an Amicon Millipore 6 ml 30-kDa MWCO centrifugal device (Amicon) and applied to a GE Superdex200 10/300 GL size exclusion column (GE Healthcare) pre-equilibrated in size exclusion buffer. Peak fractions containing stoichiometric TFIIIF were pooled, concentrated to 5–6 mg ml<sup>-1</sup> using a Vivaspin 500 MWCO 30-kDa centrifugal device (GE Healthcare), aliquots were flash-cooled in liquid nitrogen and stored at –80 °C. Typical yields were in the range of 0.1–0.2 mg per litre of *E. coli* culture.

### Preparation of the PIC

The protocol for preparation of closed mammalian PIC was adapted from the previously reported assembly schemes for the *S. cerevisiae* PIC and PIC–core Mediator complexes<sup>6</sup>. The PIC was reconstituted on a 106 nucleotide nucleic acid scaffold that had been generated from single stranded DNA oligonucleotides comprising the sequence of the AdMLP: template: 5′-AGGGAGTACTACCCCAACAGCTGGCC CTCGCAGACAGCGATGCGGAAGAGAGTGGAGGACGAACGCGCCCCA CCCCCCTTTATAGCCCCCTTCAGGAACACCCG-3′; non-template: 5′-CGGGTGTCTCTGAAGGGGGGCTATAAAAGGGGGTGGGGGCGCGTT CGTCTCACTCTCTCCGCATCGCTGTCTGCGAGGGCCAGCTGTTGG GGTGAGTACTCCCT-3′.

DNA oligonucleotides were purchased as PAGE-purified Ultramers from IDT, resuspended in buffer (100 mM KCl, 20 mM HEPES/KOH pH 7.5, 5 mM MgCl<sub>2</sub> and 2.5% (v/v) glycerol) to a concentration of 200 mM and stored at –20 °C. To anneal the strands, template and respective non-template oligonucleotides were mixed in equimolar ratios, incubated at 95 °C for 5 min and further incubated while the temperature was decreased in steps of 1 °C per 30 s to a final 4 °C. Complete 10-subunit TFIIH was reconstituted from the 7-subunit core and the 3-subunit kinase modules before PIC formation. Components of the upstream complex (DNA–TFIIA–TFIIB–TBP) and a Pol II–TFIIF complex were mixed, shortly pre-incubated in separate samples and then combined to form a Pol II–TFIIA–TFIIB–TBP–TFIIF–DNA complex. TFIIIE was incubated with previously assembled 10-subunit TFIIH for several minutes before being added to the Pol II-containing complex. Immediately, buffer S (20 mM HEPES/KOH pH 7.5, 2 mM MgCl<sub>2</sub>, 2.5% glycerol (v/v), 1 mM TCEP) with an appropriate amount of the transition state ATP-analogue ADP–BeF<sub>3</sub> was added to reach final concentrations of 1 mM ADP–BeF<sub>3</sub> and 100 mM KCl in the sample. The PIC was incubated for 120 min shaking gently at 400 rpm. All incubation steps were performed at 25 °C. The scheme for PIC reconstitution in absence of ADP–BeF<sub>3</sub> was identical but ADP–BeF<sub>3</sub> was omitted from buffer S. The purity of the ADP used for our studies was probed using phosphorylation assays, and it was found that ADP could not be used as a phosphate donor and thus is not contaminated by ATP (see Supplementary Information, Peer Review File).

The sample was centrifuged at 21,000g for 10 min and then further purified by sucrose-gradient ultracentrifugation with simultaneous

cross-linking<sup>40</sup>. The gradient was generated from equal volumes of a degassed 15% sucrose solution (15% (w/v) sucrose, 100 mM KCl, 20 mM HEPES/KOH, pH 7.5, 5 mM MgCl<sub>2</sub>, 2.5% glycerol (v/v), 1 mM TCEP, 1 mM ADP·BeF<sub>3</sub>) and a degassed 40% sucrose solution (40% (w/v) sucrose, 100 mM KCl, 20 mM HEPES/KOH, pH 7.5, 5 mM MgCl<sub>2</sub>, 2.5% glycerol (v/v), 1 mM TCEP, 1 mM ADP·BeF<sub>3</sub>, 0.2% (v/v) glutaraldehyde) with a BioComp Gradient Master 108 (BioComp Instruments). Ultracentrifugation was performed at 175,000g for 16 h at 4 °C. Subsequently, 200 µl fractions were collected and quenched with a mix of 25 mM aspartate and 40 mM lysine for 10 min. Fractions were analysed by native PAGE and Coomassie staining. Peak fractions containing cross-linked PIC were dialysed for 8 h against buffer D<sub>C</sub> (75 mM KCl, 20 mM HEPES/KOH, pH 7.5, 5 mM MgCl<sub>2</sub>, 1% (v/v) glycerol, 1 mM TCEP) in Slide-A-Lyzer MINI Dialysis Devices (2 ml, 20-kDa MWCO) (ThermoFisher Scientific) to remove sucrose and glycerol. The dialysed samples were centrifuged at 21,000g for 10 min at 4 °C and applied to cryo-EM grids. Buffers and solutions used in formation of the sucrose-gradient for ADP·BeF<sub>3</sub>-less PIC samples were devoid of the nucleotide analogue but otherwise identical to those described above.

### Promoter-dependent in vitro transcription assay

In vitro transcription assays were based on our previously reported experiments with nuclear extracts<sup>41</sup> but adapted to the mammalian recombinant system<sup>42</sup> and further modified to increase efficiency. DNA scaffolds (non-template: 5'-GGGCGTTCCTGAAGGGGGGCTATAAAGGGGGTGGGGGCGCGTTCGTCTCACTCTCTCCGCATCGCTGTCTGCGAGGGCCAGCTGTTGGGGTGAGTACTCCCTAAGCTTGGCGTAATCATGGTCATAGCTGTTTCTGTGTGAAATTGTTATCCGCTCACAA TTCCGCCC-3', template: 5'-GGGCGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTTAGGGGAGTACTACCCCAACAGCTGGCCCTCGCAGACAGCGATGCGGAAGAGAGTGAGGACGAACGCGCCCCACCCCTTTTATAGCCCCCTTCAGGAACGCC-3') were obtained from pUC119 vectors that contained the core AdMLP sequence and an approximately 65-nucleotide pUC119 window, framed by sites for the blunt-end restriction enzyme SrfI. Plasmids were prepared in large scale using a NucleoBond PC 10000 Giga kit (MACHEREY-NAGEL), followed by restriction digest with SrfI, Polyethylene-glycol precipitation and phenol-chloroform extraction.

Initiation complexes for in vitro transcription were reconstituted on scaffold DNA as described for the preparation of PIC samples for cryo-EM analysis. All incubation steps were performed at 25 °C unless indicated otherwise. Per sample, 1.6 pmol scaffold, 1.8 pmol Pol II, TFIIE and TFIIF, 5 pmol TBP and TFIIB, 9 pmol TFIIF and TFIIA were used. Reactions were prepared to reach final assay conditions of 60 mM KCl, 3 mM HEPES/KOH pH 7.9, 20 mM Tris-HCl pH 7.9, 8 mM MgCl<sub>2</sub>, 2% (w/v) polyvinyl alcohol (PVA), 3% (v/v) glycerol, 0.5 mM 1,4-dithiothreitol, 0.5 mg ml<sup>-1</sup> bovine serum albumin (BSA) and 20 units RNase inhibitor in a sample volume of 23.8 µl. To achieve complete PIC formation, samples were incubated for 30 min before transcription was initiated with 1.2 µl of 10 mM NTP solution. Transcription was permitted to proceed for 60 min before reactions were quenched by addition of 100 µl Stop buffer (300 mM NaCl, 10 mM Tris-HCl pH 7.5, 0.5 mM EDTA) and 14 µl 10% SDS, followed by treatment with 4 µg proteinase K (New England Biolabs) for 30 min at 37 °C. Nucleic acids were precipitated at -20 °C with 70% isopropanol in presence of GlycoBlue (Thermo Fisher Scientific). Precipitated samples were resuspended in water and treated with 1 unit DNase I (New England Biolabs) for 10 min at 37 °C to remove the DNA scaffold before analysis, followed by isopropanol precipitation. Samples were resuspended in 8 µl loading buffer (7 M urea, 1× TBE), applied to urea gels (7 M urea, 1× TBE, 6% acrylamide:bis-acrylamide 19:1) and separated by denaturing gel electrophoresis (urea-PAGE) in 1× TBE buffer for 45 min at 180 V. Gels were stained for 30 min with SYBR Gold (Thermo Fisher Scientific) and RNA products were visualized with a Typhoon 9500 FLA imager (GE Healthcare Life Sciences).

### Cryo-electron microscopy

For cryo-EM analysis, continuous carbon of approximately 2.8 nm thickness was floated on 130 µl of PIC sample and incubated for 4–7 min. The floated carbon was fished onto a holey-carbon grid (Quantifoil R3.5/1, copper, mesh 200), washed once with dialysis buffer and placed in a controlled environment of 100% humidity and 4 °C temperature using a Vitrobot MKIV (FEI/ThermoFischer). The grids were immediately blotted for 0.5–2 s, then plunge-frozen in liquid ethane. The data were collected automatically using SerialEM on a FEI Titan Krios (FEI/ThermoFischer) transmission electron microscope operated at 300 kV, using C2 aperture of 70 µm; slit width of 20 eV on a GIF BioQuantum energy filter (Gatan). A K3 Summit detector (Gatan) was used at a calibrated pixel size of 1.05 Å (nominal magnification of 81,000×) with a dose of approximately 40 electrons Å<sup>-2</sup> fractionated over 40 frames. A defocus range of 0.2 to 3.5 µm was used.

### Data processing and analysis

For detailed breakdown of exact micrograph and particle numbers, see Extended Data Figs. 2 and 3. The strategy used to obtain the reconstructions presented in this study is as follows. All images were initially corrected for beam and stage-induced motion by Warp and contrast transfer function (CTF) parameters were estimated using GCTF<sup>43</sup>. Motion-corrected micrograph averages were inspected manually in real-space to identify images that contained notable contaminations (ice/ethane) or uncorrectable drift and removed from the dataset. The images were then analysed in reciprocal space, and images that had notable astigmatism, poorly fitting CTF parameters, or out-of-range defoci values were also excluded from the dataset. Particles were picked from the remaining good images using Warp<sup>44</sup> and all subsequent steps were conducted using RELION-3.0.7<sup>45</sup>.

Particles were extracted with a binning factor of 5 and subjected to several rounds of reference-free 2D classification to remove falsely picked particles. The particles retained after 2D classification were re-extracted without binning and were aligned into a consensus reconstruction using 3D auto-refinement, in which the first 3D reference used was a 60 Å low-pass filtered map of the yeast PIC (EMD-3846). On the basis of these angles, 3D classification with local angular sampling was conducted to separate cPIC species (CC, OC, and particles lacking a clamp—that is, clampless). These three species of cPIC were separately subjected to CTF refinement and Bayesian polishing (three rounds each) to improve the signal within the image.

After particle polishing, cPIC classes of interest (CC for the nucleotide free, CC-like and OC for the ADP·BeF<sub>3</sub> bound datasets) were subjected to focused refinement around either cPIC (excluding the upstream complex) or the upstream complex alone to obtain maps of high-resolution. In either dataset, there were populations of particles that adopted alternative states (that is, OC-like in the dataset without ADP·BeF<sub>3</sub> and CC-like in the dataset with ADP·BeF<sub>3</sub>). These probably reflect the highly dynamic nature of the PIC. For example, the OC may be spontaneously converted back to the CC in some instances. As the flexibility of the upstream complex indicated several existing conformations of DNA within CC and OC, we classified these states by signal subtracting around the DNA and the Pol II cleft, followed by focused 3D classification without image alignment. From this, we selected classes of interest (proximal and distal CC for nucleotide-free, and the intermediary complex and OC for the ADP·BeF<sub>3</sub> bound dataset) and reverted the signal for cPIC and obtain 3D reconstructions of these cPICs. To demonstrate that the differences in DNA density between intermediary complex and distal CC were not due to a local resolution loss, local resolution analysis was conducted (Supplementary Fig. 3)

To obtain a high-resolution reconstruction of TFIIF, all polished particles of the ADP·BeF<sub>3</sub>-containing and the nucleotide-free dataset were signal subtracted of their respective cPIC, and then subjected to local re-alignment with a generous mask around where TFIIF was



anticipated to be present. This consensus refinement of all polished particles was then classified further using 3D classification with local angular sampling to identify subsets of particles that contain good signal for TFIID. These particles were then pooled and subjected to 3D auto-refinement, to obtain a reconstruction of TFIID as a whole. Focused refinement was then implemented to improve the resolution of the XPB-containing part (XPB, p8, p52) and the region encompassing the remaining chains (XPD, p62, p44, p34, MAT1) separately. During this procedure, we noticed that occasionally density for the MAT1 RING domain was lacking, and so we classified particles that were pre-aligned onto the XPD-containing part and conducted 3D classification without image alignment to enrich for particles that showed the most promising density for the MAT1 RING finger domain (which indicates that TFIID is attached to the stalk). Further 3D classification based on cPIC state did not reveal any conformation or compositional differences in TFIID. Thus, we applied each consensus TFIID reconstruction as the map for all cPIC states within each dataset.

Finally, to reconstruct the entire PIC, we identified particles present in the respective cPIC and TFIID subsets of data (that is, a certain cPIC state with attached TFIID) and these were subjected to 3D auto-refinement. Even after this extensive classification procedure that ensured that particles contained both cPIC and TFIID, conformational heterogeneity outstanding precluded a straightforward visualization of cPIC and TFIID in the same reconstruction as the signal of cPIC dominated the alignment and resulted in very diffuse density for TFIID in an unmasked refinement. To circumvent this, we applied a non-binary mask, in which areas of strong signal (that is, cPIC) were attenuated for the purposes of angular assignment only (that is, the final output is not directly affected) by a mask with values of less than one which resulted in a more balanced reconstruction where both cPIC and TFIID have similar local resolutions, albeit at a lower overall resolution. Analysis of the conformational relationship between cPIC and TFIID were further probed using multibody analysis and 3D classification, and demonstrated that the motions involved were distributed as a Gaussian and no distinct alternative conformational energetic minima were detected (Supplementary Fig. 4).

Resolution was estimated using a Fourier shell correlation cut-off of 0.143 between the two reconstructed half maps. Before using the reconstructions for model building, the reconstructions were filtered by their local resolution using RELION, and map-sharpening was conducted using phenix.autosharp.

## Model building and refinement

For model building, previously published structures (PDB codes 5FLM, 5IY6, 5IY7, 5GPY, 6RO4 and 6NMI) were used either as starting templates or rigid bodies for docking into our cryo-EM density. Where the cryo-EM map resolution permitted, the model was inspected residue-by-residue and modified to improve the fit to the map while maintain favourable geometry. In places where our resolution approached 4.0 Å (for example, XPD of TFIID), we limited the building process mainly to rigid body docking of previous structures and resolving clashes. Iterative rounds of manual rebuilding in Coot<sup>46</sup> together with real space refinement in PHENIX<sup>47</sup> resulted in final models with good geometry. All structural figures were generated using PyMol.

## Reporting summary

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

## Data availability

The cryo-EM density reconstructions and final models have been deposited with the Electron Microscopy Data Bank (EMDB) (accession codes EMD-12611–EMD-12630) and with the Protein Data Bank (PDB) (proximal cPIC: 7NVS, distal cPIC: 7NVT, TFIID (CC): 7NVW, cPIC (OC): 7NVU, TFIID (OC): 7NVX, XPB (intermediary complex): 7NVV, proximal CC: 7NVY, distal CC: 7NVZ, OC: 7NW0). All data are available in the Article or its Supplementary Information files. Publicly available PDB entries used in this study are: 5FLM, 5IY7, 5GPY, 6NMI and 6RO4.

35. Gradia, S. D. et al. MacroBac: new technologies for robust and efficient large-scale production of recombinant multiprotein complexes. *Methods Enzymol.* **592**, 1–26 (2017).
36. Berrow, N. S. et al. A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Res.* **35**, e45 (2007).
37. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).
38. Vos, S. M., Farnung, L., Urlaub, H. & Cramer, P. Structure of paused transcription complex Pol II–DSIF–NELF. *Nature* **560**, 601–606 (2018).
39. Hu, X. et al. A Mediator-responsive form of metazoan RNA polymerase II. *Proc. Natl Acad. Sci. USA* **103**, 9506–9511 (2006).
40. Kastner, B. et al. GraFix: sample preparation for single-particle electron cryomicroscopy. *Nat. Methods* **5**, 53–55 (2008).
41. Seitz, M., Larivière, L., Pfaffeneder, T., Wenzel, L. & Cramer, P. Mediator head subcomplex Med11/22 contains a common helix bundle building block with a specific function in transcription initiation complex stabilization. *Nucleic Acids Res.* **39**, 6291–6304 (2011).
42. Noe Gonzalez, M., Sato, S., Tomomori-Sato, C., Conaway, J. W. & Conaway, R. C. CTD-dependent and -independent mechanisms govern co-transcriptional capping of Pol II transcripts. *Nat. Commun.* **9**, 3392 (2018).
43. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
44. Tegunov, D. & Cramer, P. Real-time cryo-electron microscopy data preprocessing with Warp. *Nat. Methods* **16**, 1146–1152 (2019).
45. Scheres, S. H. W. A Bayesian view on cryo-EM structure determination. *J. Mol. Biol.* **415**, 406–418 (2012).
46. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
47. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D* **75**, 861–877 (2019).

**Acknowledgements** We thank members of the Cramer laboratory, in particular C. Dienemann and U. Steuerwald for maintenance of the cryo-EM facility, U. Neef and P. Rus for maintenance of the insect cell facility, and F. Jochheim and M. Klein for maintenance of the computing infrastructure. S.A. was supported by an H2020 Marie Curie Individual Fellowship (894862). P.C. was supported by the Deutsche Forschungsgemeinschaft (EXC 2067/1 39072994, SFB860, SPP2191) and the ERC Advanced Investigator Grant CHROMATRANS (grant agreement no. 882357).

**Author contributions** S.A. designed and carried out data analysis and model building. S.S. designed and carried out experiments. P.C. designed and supervised research. S.A., S.S. and P.C. interpreted the data and wrote the manuscript.

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

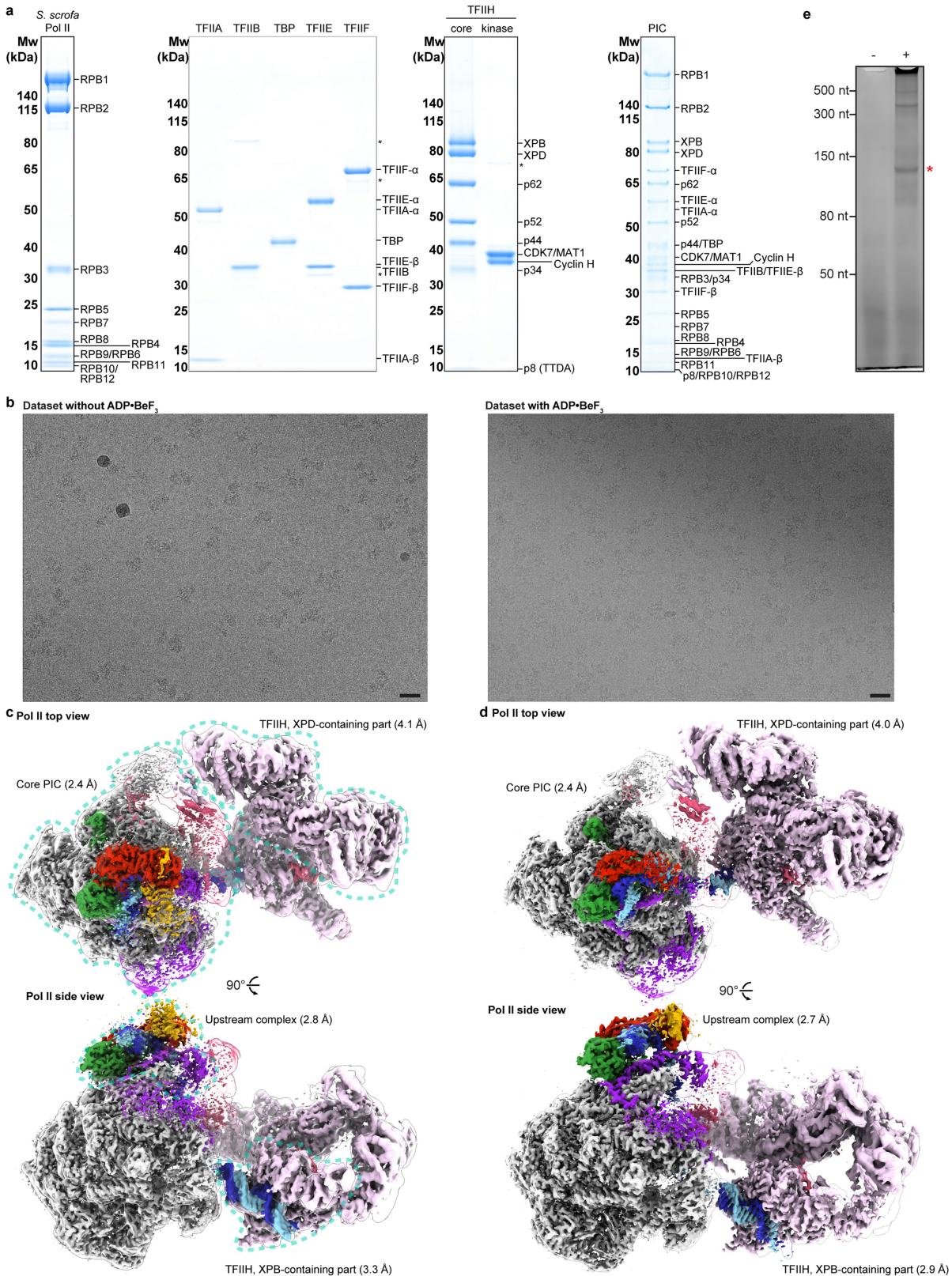
## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41586-021-03554-8>.

**Correspondence and requests for materials** should be addressed to P.C.

**Peer review information** Nature thanks Steve Hahn and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

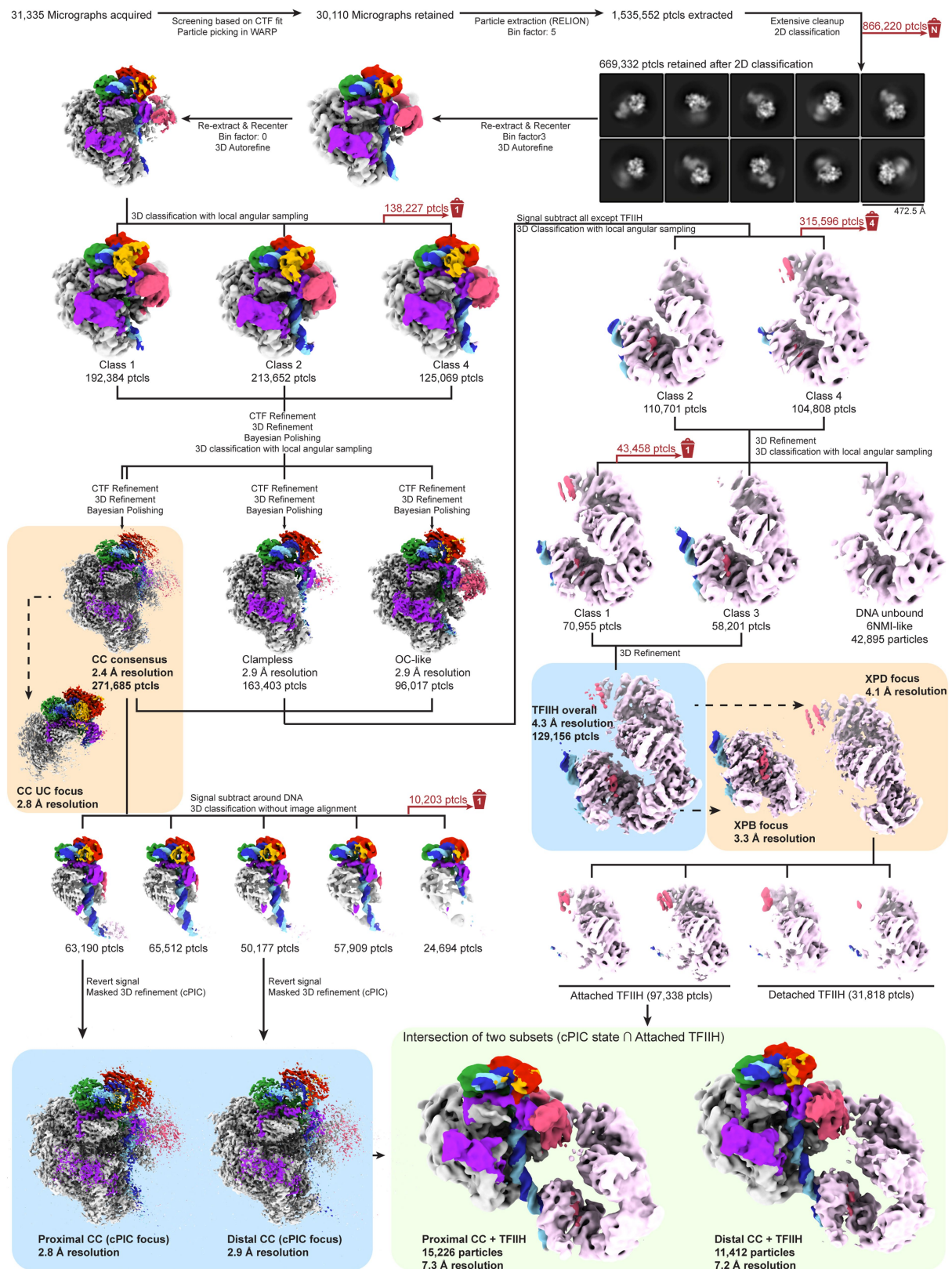
**Reprints and permissions information** is available at <http://www.nature.com/reprints>.



**Extended Data Fig. 1 | Biochemical assembly of PIC samples, and PIC cryo-EM maps. a**, SDS-PAGE analysis of *S. scrofa domestica* Pol II, *H. sapiens* general transcription factors and a peak fraction of mammalian PIC assembled on a sucrose gradient (replicated three times). **b**, Representative cryo-EM micrograph. Left, without ADP·BeF<sub>3</sub> (replicated 30,110 times). Right, with ADP·BeF<sub>3</sub> (replicated 67,108 times). Scale bar, 300 Å. **c, d**, Cryo-EM reconstructions used for model building of the CC-states (without ADP·BeF<sub>3</sub>)

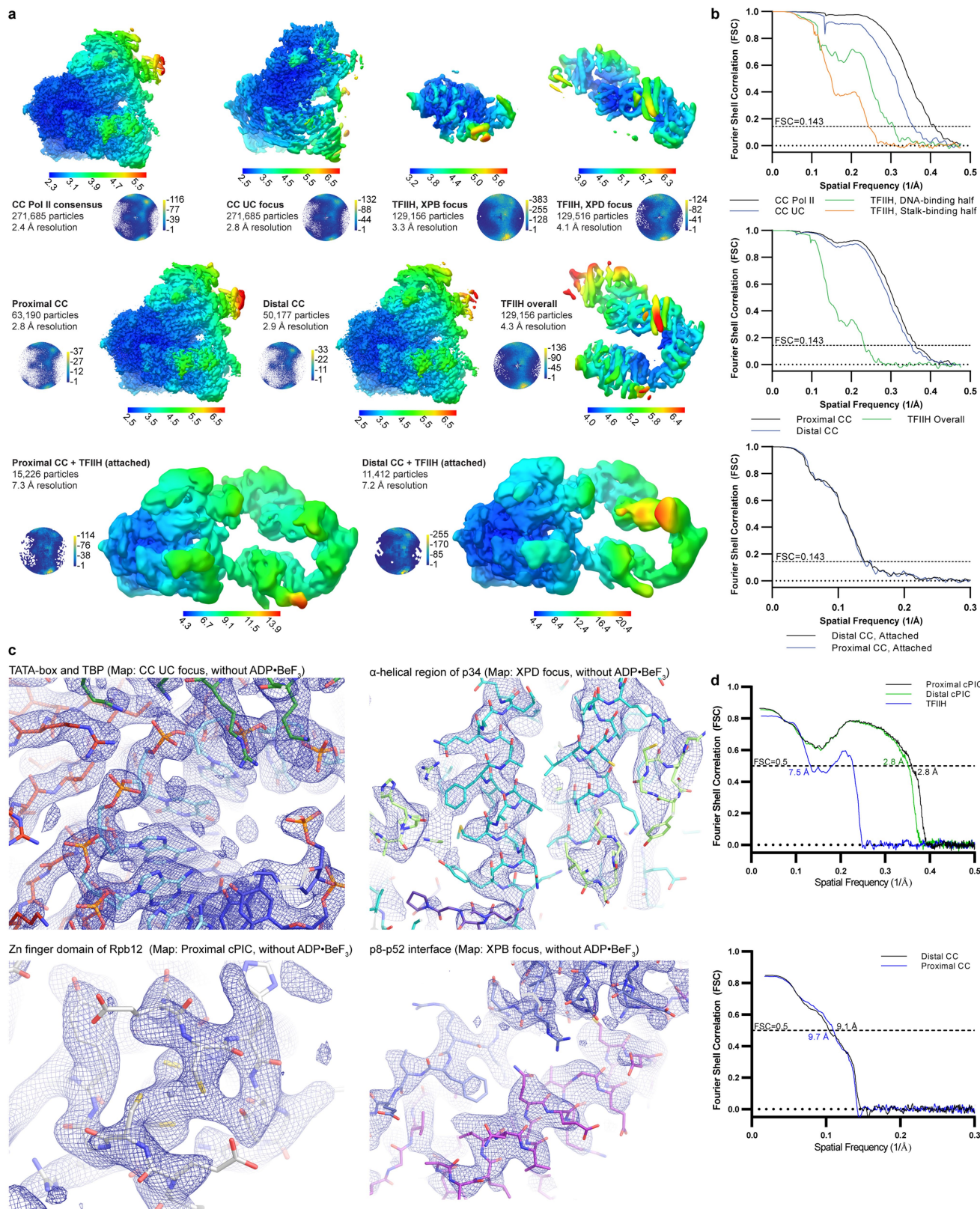
(c) and OC-state (with ADP·BeF<sub>3</sub>) (d). The resolution for the respective focused refined maps is provided in parentheses, and an outline of the lower-resolution map used to dock in the focused maps is shown. **e**, Promoter-dependent transcription assay demonstrates that our PIC preparation is initiation competent. Asterisk indicates expected RNA product length. For the negative control, TFIIE and TFIIF were lacking from the reaction (replicated three times).





**Extended Data Fig. 2 | Processing tree of cryo-EM dataset of PIC without ADP·BeF<sub>3</sub>.** Processing tree indicating how the particles were assigned their classes and angles for this study. Reconstructions with an orange backing indicate high-resolution maps used for initial building, which were then fitted into reconstructions with blue background that encompassed either the entire cPIC or TFIIH, and these maps were fitting into the reconstructions with a green

background that indicate the reconstructions of the entire PIC. Where classes were discarded due to poor quality during classification, the number of particles and the number of classes they encompassed are given as a red bin (number on the bin indicates the number of classes discarded in a given classification run).



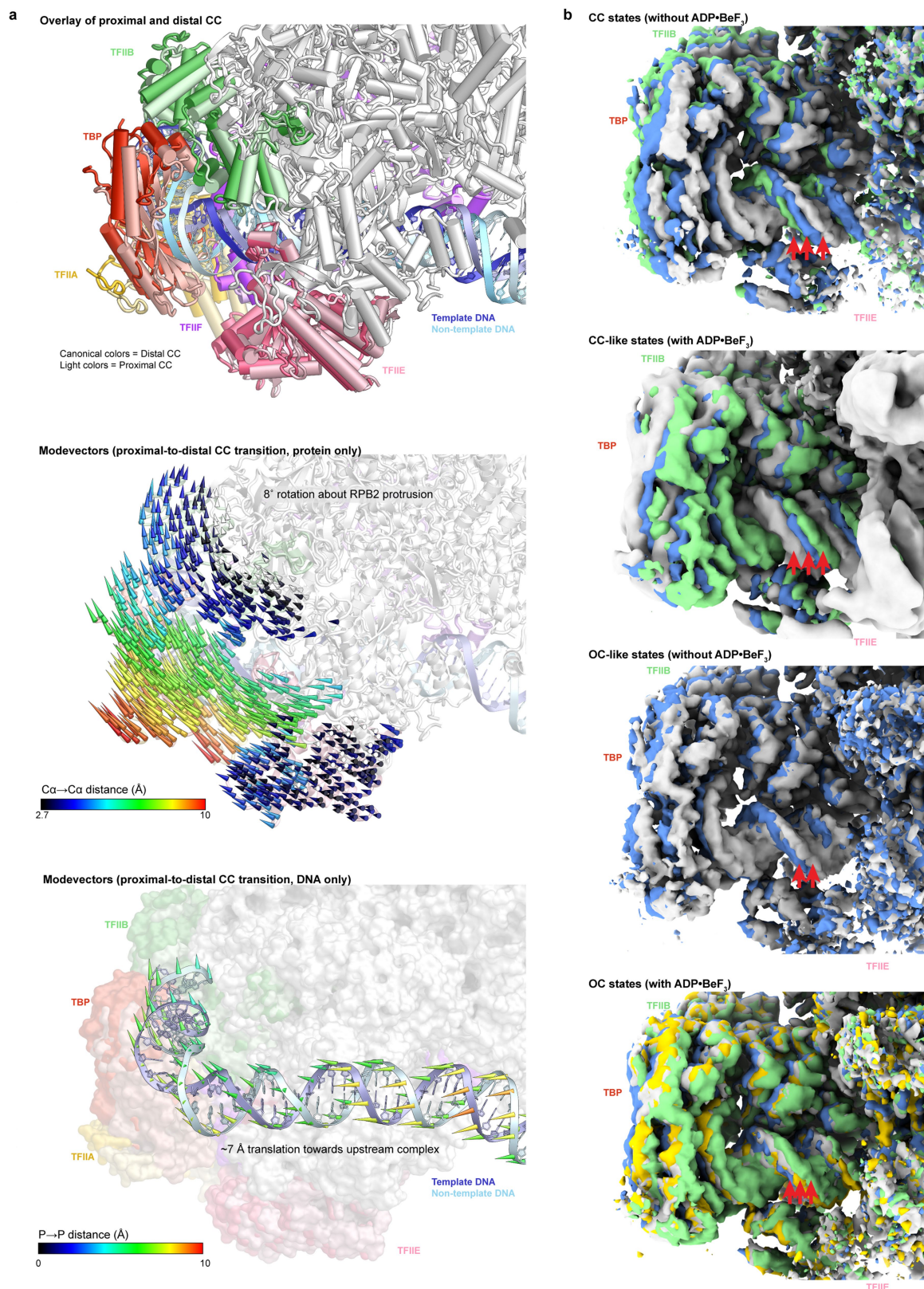
**Extended Data Fig. 3 | Local resolution distribution, FSCs and representative cryo-EM density for the dataset without ADP·BeF<sub>3</sub>.**

**a**, Reconstructions obtained from the dataset without ADP·BeF<sub>3</sub> coloured by their local resolution as estimated using RELION. Total particle count, the global resolution estimate using the Fourier shell correlation (FSC) = 0.143 criterion are given together with an angular distribution plot.

**b**, Solvent-corrected 'gold-standard' FSCs grouped according to the

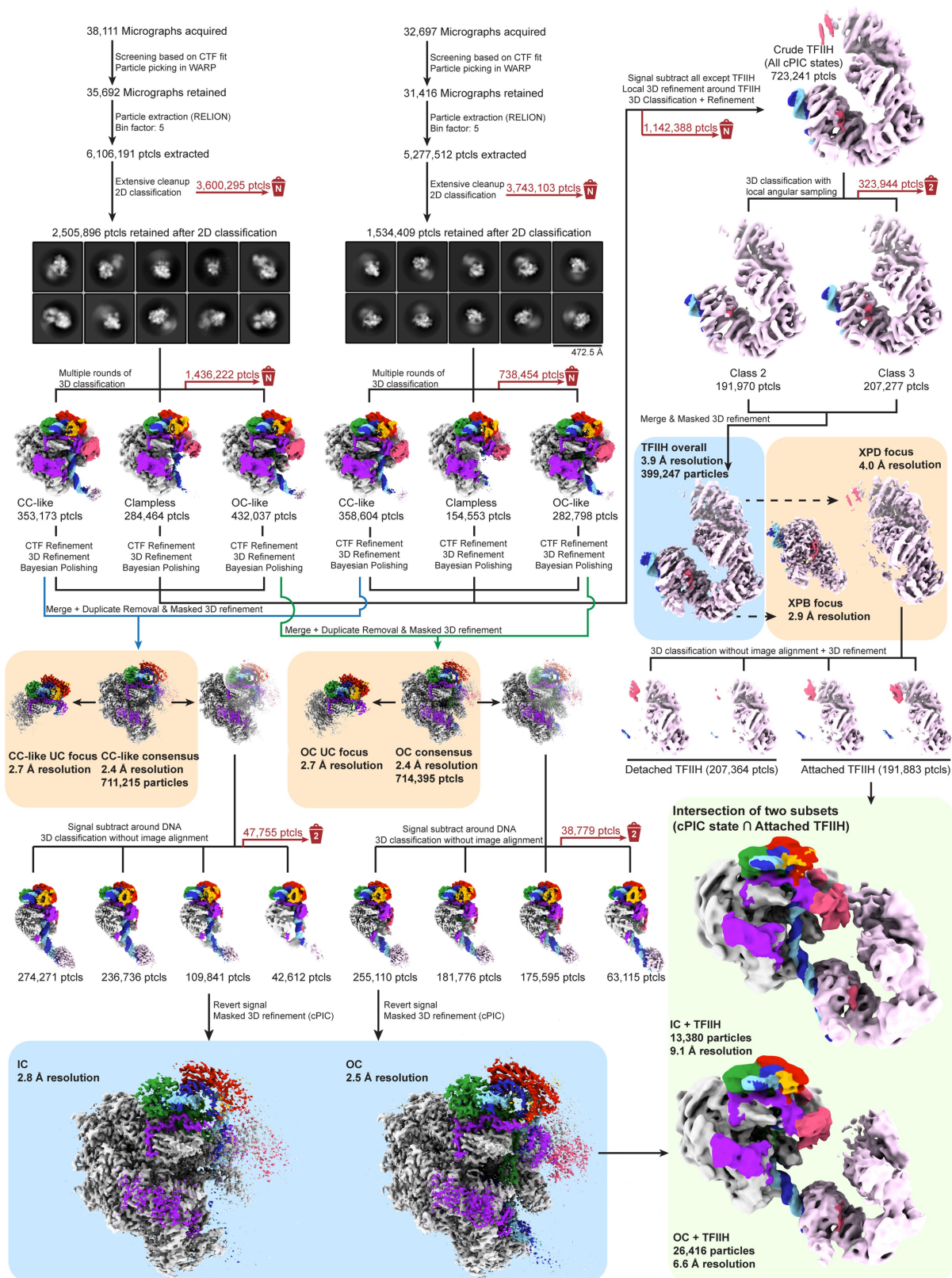
reconstructions shown in **a**. **c**, Exemplary cryo-EM density (shown as blue mesh) with the fitted atomic models shown with their canonical chain colours. The particular reconstruction and model displayed are specified in the heading of each image. **d**, Model-to-map FSCs. FSCs (cut-off = 0.5) are grouped according to their resolution-tier going from most focused maps on the top to the overall maps on the bottom.





**Extended Data Fig. 4 | Upstream flexibility causes variable DNA positions within the cPIC. a,** Structural overlay of the two CC states observed within the PIC dataset without ADP•BeF<sub>3</sub> on RPB1. All cPIC components align well except for the upstream complex. Analysis of the difference between these two observed states shows that the upstream complex rotates about the RPB2 protrusion element, and the DNA translates about 7 Å laterally across the cleft. No notable rotation in the DNA was observed, which indicates that this motion

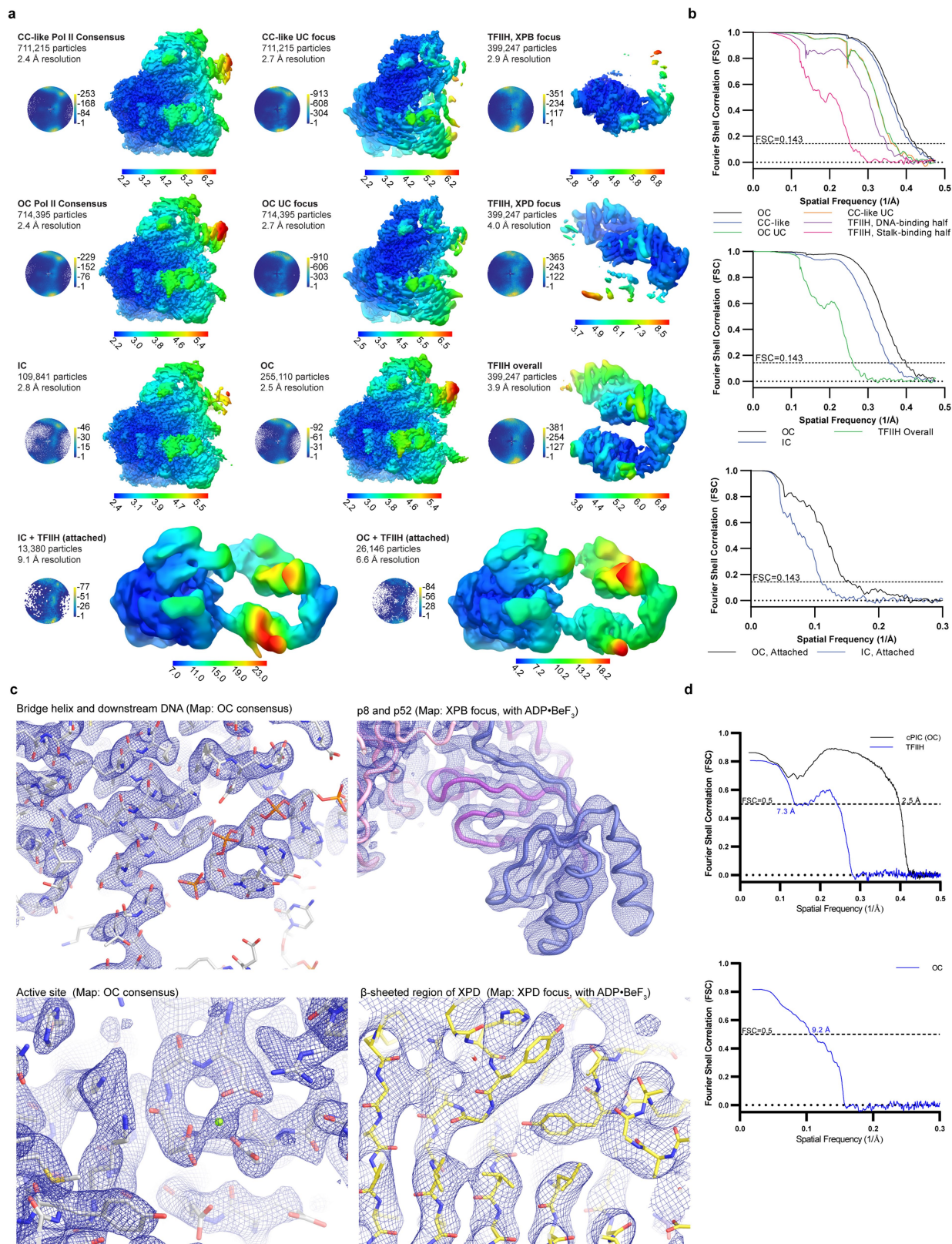
is independent of translocation activity. **b,** Cryo-EM density showing the placement of DNA in the distinct upstream complex classes when the separate classes are aligned on the Pol II core. Red arrows indicate positions of the phosphate backbone in the respective classes. It is notable that in OC-like classes, the motion is of a lesser degree than in the CC classes. The position of TBP is denoted for orientation.



**Extended Data Fig. 5 | Processing tree of the cryo-EM dataset of PIC with ADP·BeF<sub>3</sub>.** Processing tree indicating how the particles were assigned their classes and angles for this study. Reconstructions with an orange backing indicate high-resolution maps used for initial building, which were then fitted into reconstructions with a blue background that encompassed either the entire cPIC or TFIIH, and these maps were fitted into the reconstructions with a

green background that indicate the reconstructions of the entire PIC. Where classes were discarded due to poor quality during classification, the number of particles and the number of classes they encompassed are given as a red bin (number on the bin indicates the number of classes discarded in a given classification run).





**Extended Data Fig. 6 | Local resolution distribution, FSCs and representative cryo-EM density for the dataset with ADP•BeF<sub>3</sub>.**

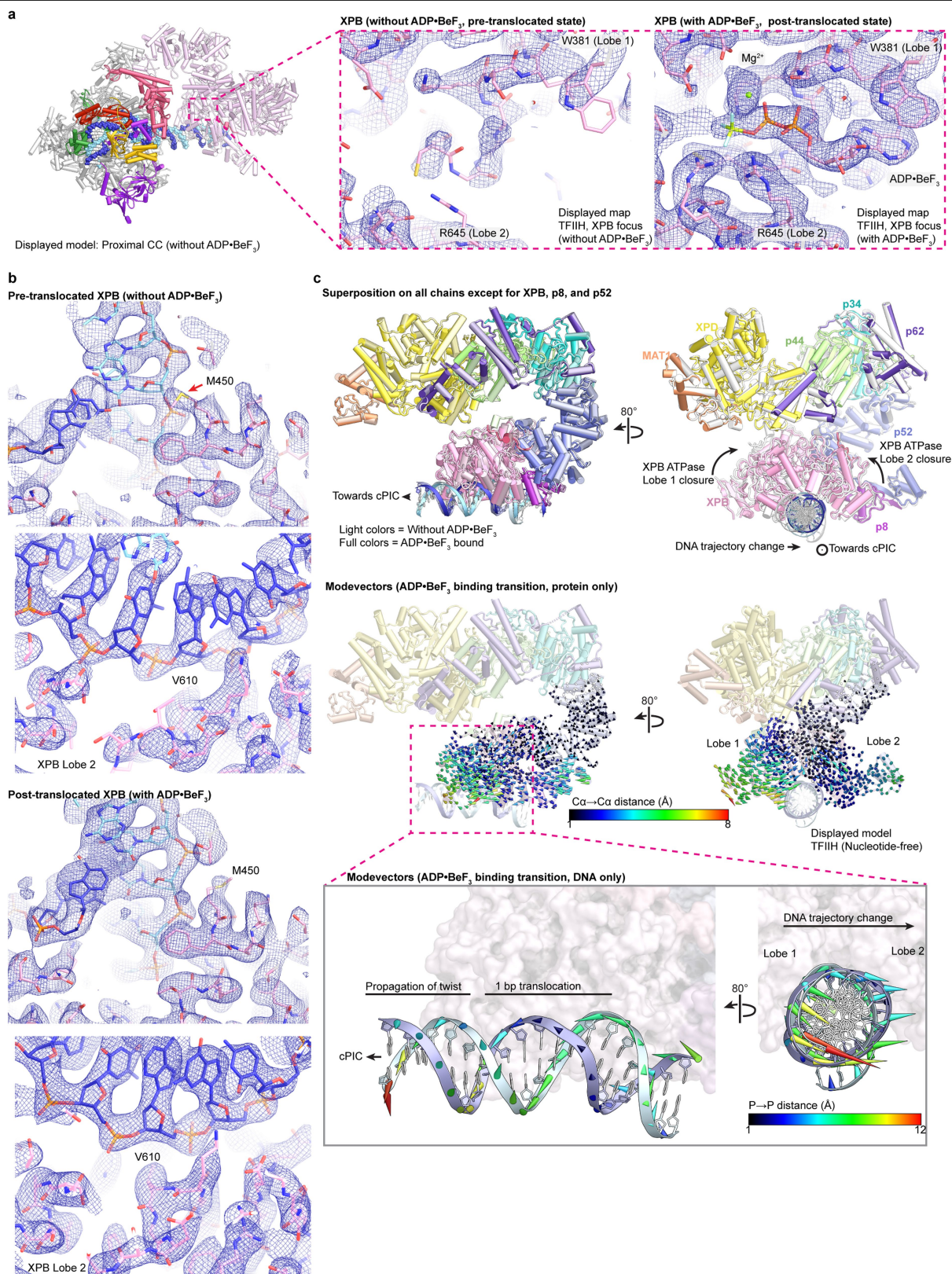
**a**, Reconstructions obtained from the dataset with ADP•BeF<sub>3</sub> coloured by their local resolution as estimated using RELION. Total particle count, the global resolution estimate using the FSC = 0.143 criterion are given together with an angular distribution plot. **b**, Solvent-corrected 'gold-standard' FSCs grouped

according to the reconstructions shown in **a**. **c**, Exemplary cryo-EM density (shown as blue mesh) with the fitted atomic models shown with their canonical chain colours. The particular reconstruction and model displayed are specified in the heading of each image. **d**, Model-to-map FSCs. FSCs (cut-off = 0.5) are grouped according to their resolution-tier going from most focused maps on the top to the overall maps on the bottom.









**Extended Data Fig. 8 | TFIIF conformational change after ADP·BeF<sub>3</sub> binding.** **a**, Detailed view into the ATP-binding pocket of XPB. Our cryo-EM maps allowed for the unambiguous distinction between the unoccupied pocketed (left, without ADP·BeF<sub>3</sub>) and the ADP·BeF<sub>3</sub>-bound state with a coordinating magnesium ion. **b**, Side-by-side comparison of the interactions between XPB and DNA in pre- and post-translocated XPB. Essentially all interactions between XPB and the major or minor grooves are unchanged with the exception of M450 which appears to flip out of the minor groove in the

post-translocated state. **c**, Superposition of the pre- and post-translocated TFIIF on all chains except for XPB, p8 and p52 show the conformational changes in the context of the PIC. Modevectors (right) show that both ATPase lobes of XPB close onto the DNA after ADP·BeF<sub>3</sub> binding with lobe 1 moving more than lobe 2. After translocation, the DNA twist stored within XPB is propagated towards cPIC and at the same time a change in trajectory is observed.







## Extended Data Fig. 9 | Pol II clamp closure and dissociation of the MAT1 RING domain.

**a**, The closed cPIC viewed from a rotated top view across the closed promoter DNA. The initiator (INR) is found flanked by the Pol II clamp head and the TFIIIF charged helix. We observe a partial unwinding of this region after a power-stroke by XPB introduced torsional strain in this area. For clarity, both maps were filtered to a resolution of 5 Å. IC, intermediary complex.

**b**, Conformational change that occurs during the transition between CC and OC. The clamp of Pol II closes upon the DNA, rotating approximately 11° about the Pol II core. Together with the clamp, the stalk of Pol II together with TFIIIE buckle inwards. Although a closed clamp is not compatible with closed DNA owing to steric clashes, an open clamp does not clash with open DNA, but is not observed in our data. The CC-to-OC transition in the human PIC also involves restructuring of fork loop 2 from a compact helical conformation to an

extended conformation that is not found in the yeast OC (not shown). **c**, Overview of TFIIIB that underwent substantial remodelling. Our model is coloured green, and yeast TFIIIB is shown as an outline, and missing secondary structural elements compared with yeast are indicated. Density-to-model fits in four areas of TFIIIB (B-ribbon, B-reader, B-linker and B-core), demonstrate excellent agreement of the model to the experimental data. **d**, Detailed view of the proteinaceous interface between TFIIH and Pol II. The MAT1 RING finger domain nestles between the RPB7 portion of the stalk and on top of the TFIIIE 'E-linker' helices. Clear cryo-EM density can be observed for the MAT1 RING finger domain in all classes except for the OC class. This indicates that this Pol II-TFIIH interface is disrupted during OC formation but not a direct consequence of XPB translocation, as the intermediary complex class that is in a post-translocated state still shows density for the MAT1 RING finger domain.

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☐ ☒ The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☒ ☐ The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- ☒ ☐ A description of all covariates tested
- ☒ ☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☒ ☐ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☒ ☐ For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

### Software and code

Policy information about [availability of computer code](#)

Data collection SerialEM 3.8

Data analysis RELION 3.0.7, Warp1.0.7, GCTF 1.06, Coot 0.8.9.3, PHENIX-1.18, GraphPad Prism 8, Chimera 1.13.1, ChimeraX 1.1, Pymol-2.3.3, Molprobit 4.5.1

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The cryo-EM density reconstructions and final models were deposited with the EMDB (entries EMD-12611-12630) and with the PDB (Proximal cPIC: 7NVS, Distal cPIC: 7NVT, TFIH (CC): 7NVW, cPIC (OC): 7NVU, TFIH (OC): 7NVX, XPB (IC): 7NVV, Proximal CC: 7NVY, Distal CC: 7NVZ, OC: 7NW0). All data is available in the main text or the supplementary materials. Publicly available PDB entries used in this study: 5FLM, 5IY7, 5GPY, 6NMI, 6RO4

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to predetermine sample size. Structural data was collected on two independently prepared samples. As many independently recorded images are acquired as part of cryo-EM data collection, the sample sizes are sufficient.
Data exclusions	No data were excluded from the analyses. However, as part of the analysis data were clustered and some clusters were processed further than others
Replication	All attempts at replication were successful. Cryo-EM reconstructions inherently contains a high degree of multiplicity arising from being averaged over a large number of independent observations. During the processing pipeline, replicate reconstructions were calculated over 3 times during the polishing and other related refinement procedures.
Randomization	Samples were not allocated to experimental groups, as this is not a procedure relevant in cryo-EM data processing
Blinding	Investigators were not blinded during data acquisition and analysis because it is not a common procedure for the methods employed. The methods would not be possible if the investigators were blinded, and furthermore the data processing procedure requires a priori knowledge of the sample to be performed optimally.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

### Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Antibodies

Antibodies used	8WG16 Antibody (purification of Pol II, purified in-house and conjugated to resin), dilution: N/A
Validation	SDS-PAGE analysis of the resulting purified material as well as high resolution structure determination by cryo-EM

## Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	Hi5 cells: Expression Systems, Tni Insect Cells in ESF921 media, Item 94-002F Sf9 cells: ThermoFisher, Catalogue Number 12659017, Sf9 cells in Sf-900TM III SFM
Authentication	Isozyme and karyotype analysis by commercial supplier (ThermoFisher and Expression Systems). Cell lines were not authenticated in-house.
Mycoplasma contamination	Mycoplasma test was not required for used cell lines.
Commonly misidentified lines (See <a href="#">ICLAC</a> register)	No commonly misidentified cell lines were used