ARTICLE

doi:10.1038/nature17394

Structure of promoter-bound TFIID and model of human pre-initiation complex assembly

Robert K. Louder[1], Yuan He[2,3]†, José Ramón López-Blanco[4], Jie Fang[5], Pablo Chacón[4] & Eva Nogales[2,3,5]

The general transcription factor IID (TFIID) plays a central role in the initiation of RNA polymerase II (Pol II)-dependent transcription by nucleating pre-initiation complex (PIC) assembly at the core promoter. TFIID comprises the TATA-binding protein (TBP) and 13 TBP-associated factors (TAF1–13), which specifically interact with a variety of core promoter DNA sequences. Here we present the structure of human TFIID in complex with TFIIA and core promoter DNA, determined by single-particle cryo-electron microscopy at sub-nanometre resolution. All core promoter elements are contacted by subunits of TFIID, with TAF1 and TAF2 mediating major interactions with the downstream promoter. TFIIA bridges the TBP–TATA complex with lobe B of TFIID. We also present the cryo-electron microscopy reconstruction of a fully assembled human TAF-less PIC. Superposition of common elements between the two structures provides novel insights into the general role of TFIID in promoter recognition, PIC assembly, and transcription initiation.

Initiation of transcription by Pol II represents a major control point for eukaryotic cells, and its regulation is the primary means of differential gene expression in metazoans[1]. A prerequisite for Pol II transcription is the recruitment of the general transcription factors TFIIA, -B, -D, -E, -F, and -H, to the core promoter, where they assemble with Pol II into the PIC[2]. The process is thought to begin with the recruitment of TFIID and TFIIA to the core promoter, followed by TFIIB, TFIIF, and Pol II, and ending with TFIIE and TFIIH[2,3].

TFIID is an ~1 megadalton complex consisting of TBP and TAF1–13 (refs 4, 5). TBP and TAFs mediate specific interactions with a variety of core promoter sequences[6–10] and other components of the PIC[11–14], establishing TFIID as the primary core promoter recognition factor that nucleates PIC assembly. Additionally, TAFs mediate regulatory signals by interacting with transcriptional activators or epigenetic marks, and TFIID has been shown to be required for the initiation of activated transcription[5].

Despite its critical role in transcription, little is known about the arrangement of subunits within TFIID and the structural bases of their interactions with DNA and the transcriptional machinery. The lack of a recombinant expression system for full TFIID necessitates purification from endogenous sources, which limits the yield that can be used for structural studies. There are crystallographic structures for domains of several TFIID subunits[15–25], but only low-resolution electron microscopy (EM) structures of the TFIID holocomplex and subcomplexes[25–33].

Here we present the cryo-EM structure of human TFIID bound to TFIIA and core promoter DNA, determined by single-particle cryo-EM at 7–16 Å resolution. The structure reveals the position of the TBP–TFIIA–TATA subcomplex and defines the path and register for promoter DNA. Our study also shows the locations of TAF1, -2, -6, -7, and -8, and implicates specific elements within TAF1 and TAF2 in mediating interactions with downstream core promoter DNA. We also present the cryo-EM reconstruction of a human TAF-less PIC containing Pol II, promoter DNA, TBP, TFIIA, -B, -F, -S, -E, and -H. By superimposing the common elements between the two structures, we propose a model for the complete TFIID-based PIC that provides novel insights into the intertwined roles of TFIID in promoter recognition, PIC assembly, and transcription initiation.

Overall structure of promoter-bound TFIID

Human TFIID has a horseshoe shape, with lobes A, B, and C surrounding a central cavity[26,29]. Our previous cryo-EM studies revealed that human TFIID adopts two major conformations, termed the canonical and rearranged states, that differ in the position of lobe A[33]. In the canonical state, lobe A is attached to lobe C, while in the rearranged state, which is the conformation in which TFIID binds promoter DNA, lobe A adopts a position proximal to lobe B. To reduce the conformational and compositional heterogeneity limiting the resolution of our previous studies (~30 Å), we purified human promoter-bound TFIID complexes using the super core promoter (SCP) sequence. This composite promoter was designed to maximize transcriptional output by increasing the affinity of TFIID for DNA through the presence of several naturally occurring promoter motifs (TATA, Inr, MTE, and DPE)[34]. Purification of SCP-bound TFIID in the presence of TFIIA (see Methods) resulted in more homogeneous TFIID-IIA–SCP complexes, which we then used in single-particle cryo-EM to obtain a reconstruction with an overall resolution of 10.2 Å (Extended Data Fig. 1).

The shape of the human promoter-bound TFIID is consistent with previous lower-resolution reconstructions[26,29,33]. In our present structure, lobe A appears separated into a smaller lobe (lobe A1) that is more stably positioned with respect to the BC core, and a highly flexible lobe A2 (Fig. 1a and Extended Data Fig. 1e). To improve the resolution of the more stable core (comprising lobes A1, B, and C), lobe A2 was excluded from the references used during subsequent three-dimensional classification and refinement (see Methods). This procedure led to an improved reconstruction of the promoter-bound TFIID core with an overall resolution of 8.7 Å (Fig. 1b and Extended Data Fig. 2a, c–e).

[1]Biophysics Graduate Group, University of California, Berkeley, California 94720, USA. [2]QB3 Institute, Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA. [3]Molecular Biophysics and Integrative Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. [4]Department of Biological Physical Chemistry, Rocasolano Physical Chemistry Institute, CSIC, Serrano 119, Madrid 28006, Spain. [5]Howard Hughes Medical Institute, University of California, Berkeley, California 94720, USA.
†Present address: Interdisciplinary Biological Sciences Program, Northwestern University, Evanston, Illinois 60208, USA.

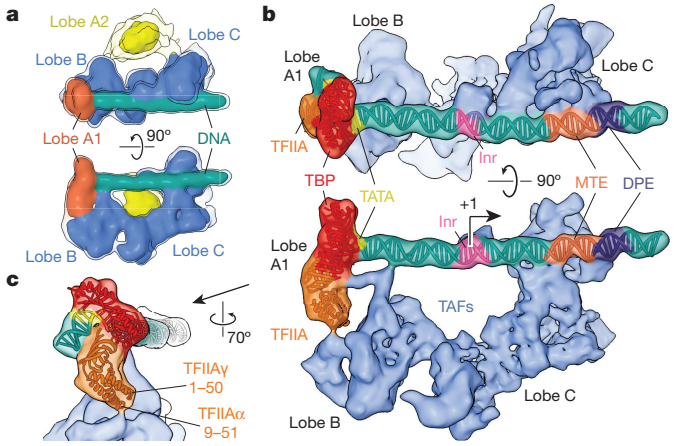


Figure 1 | Cryo-EM reconstruction of the TFIID-IIA–SCP complex. **a**, TFIID-IIA–SCP reconstruction. Isosurfaces are displayed at two thresholds, with the lower one shown in transparency to enable visualization of weaker densities. **b**, Locally refined cryo-EM reconstruction of the promoter-bound core of TFIID (that is, excluding lobe A2). TSS is marked '+1' and the transcription direction by an arrow. **c**, Close-up view of the TBP–TFIIA promoter-binding module, indicating putative TFIID-interacting regions of TFIIA.

TFIIA–TBP–TATA subcomplex and DNA path

We could easily localize the TBP–TFIIA–TATA ternary complex within lobe A1 by rigid-body docking of the crystal structure[19] into the cryo-EM density (Fig. 1c and Supplementary Video 1). This result is consistent with previous EM studies using gold labelling that localized the TATA-containing upstream promoter region to lobes A and B, and the downstream region to lobe C of the rearranged TFIID[33]. The position of TBP and TFIIA in our structure, together with previous lower-resolution reconstructions of promoter-bound TFIID-IIA and of TFIID alone[33] (Extended Data Fig. 3a), are consistent with the proposal that TBP resides in the mobile lobe A and thus changes position during TFIID rearrangement.

In our promoter-bound complex, TFIIA appears to serve as a bridge between TBP and lobe B of TFIID. Interestingly, DNase I footprinting of TFIID-bound SCP showed that only in the presence of TFIIA is the TATA box protected[33], supporting the idea that TFIIA is essential for TBP positioning for DNA engagement in the rearranged state. The tip of the TFIIA four-helix bundle is oriented towards lobe B, in agreement with the finding that mutations within this region of TFIIA affect its interaction with TFIID[35] (Fig. 1c and Extended Data Fig. 3b). A density connecting the TFIIA β-barrel with lobe B of TFIID and the DNA near position −16 (Fig. 1b, bottom, and Extended Data Fig. 3b) cannot reasonably accommodate any TFIIA portions unmodelled in the crystal structure. On the other hand, footprinting has shown a TFIID-dependent protection from DNase I cleavage in this region of promoter DNA[33], suggesting the connecting density probably corresponds to one of the TAFs in lobe B.

We could model the SCP DNA from −40 to +42 base pairs (bp) relative to the transcription start site (TSS), including all SCP motifs (TATA, Inr, MTE, and DPE), using the bent TATA DNA as an anchoring point for assigning the base pair register (Fig. 1b and Supplementary Video 1). The MPE-Fe cleavage pattern of SCP DNA bound to TFIID-IIA[33] can be mapped onto the structure with high correspondence between protected sequences and protein contacts (Extended Data Fig. 3c). Mapping the downstream core element[9] onto our structure strongly suggests that TFIID uses very similar protein elements to interact with this alternative promoter motif (Extended Data Fig. 3d).

Lobe C architecture and downstream promoter binding

The structural stability of lobe C and bound downstream promoter DNA relative to the rest of the complex allowed us to improve the resolution in this region (to ~7–12 Å, with an 8.2 Å average) through further local three-dimensional classification and alignment against a masked reference (Extended Data Fig. 2b–f). The crystal structure of a human TAF1–TAF7 complex[24] that includes the highly conserved central and amino (N)-terminal fragments of TAF1 and TAF7, respectively, could be unambiguously docked as a rigid body into the density adjacent to the downstream core promoter (Fig. 2a and Supplementary Video 1), in agreement with its reported ability to bind DNA with a preference for the downstream sequence of the SCP[24]. The docking reveals that TAF1 is the primary mediator of downstream promoter binding, contributing contacts that span 34 bp of DNA (positions −3 to +31; Fig. 2a). The winged helix (WH) domain of TAF1 forms a major interaction at the junction of the MTE and DPE promoter motifs (Fig. 2a–c). Superposition of the TAF1 WH–DNA complex with other DNA-binding WH proteins confirms that it shares a common mode of DNA recognition[36] (Extended Data Fig. 4a). The third α-helix (α3) of this domain inserts into the major groove, positioning three conserved positively charged residues (R864, K865, K868) for specific interaction with the MTE (Fig. 2b, c and Extended Data Fig. 4b), which supports the finding that mutation of these residues to alanine ablates binding of the TAF1–TAF7 module to promoter DNA[24]. Additional conserved positively charged residues within the extended β-wing (R875) and N-terminal end of the first α-helix (α1; K818) of the WH domain form additional promoter contacts within the minor groove of the DPE and near the upstream boundary of the MTE, respectively (Fig. 2b, c and Extended Data Fig. 4b). There is a small protein density contacting the minor groove of the Inr (Fig. 2a, d). We propose that it corresponds to the portion of TAF1 between residues 993–1075, which is disordered in the crystal structure but is well conserved among metazoans and is predicted to

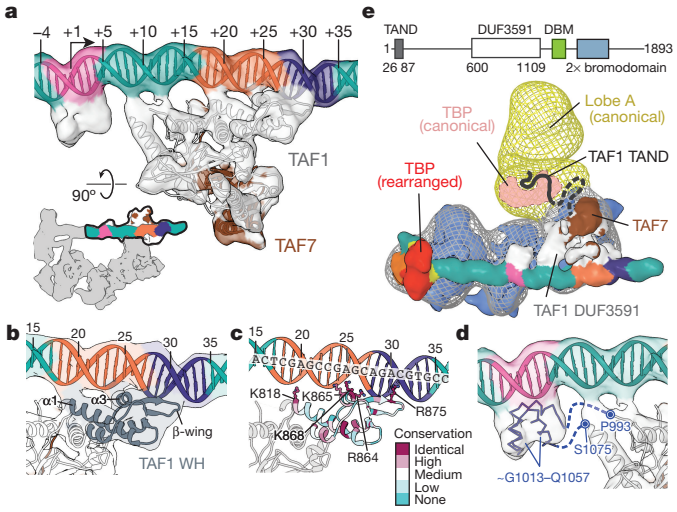


Figure 2 | A TAF1–TAF7 subcomplex forms a downstream promoter-binding module. **a**, Docking of the human TAF1–TAF7 complex (PDB accession number 4RGW)[24] into the locally refined lobe C density. Promoter is coloured as in Fig. 1. The location of the segmented density in the overall map is highlighted in the bottom left. **b**, Close-up view of the TAF1 WH domain (dark grey) bound to promoter DNA. **c**, The TAF1 WH domain with residues coloured according to conservation (Extended Data Fig. 4a). Conserved positively charged residues that appear involved in DNA binding are shown as ball-and-sticks. **d**, Predicted three-dimensional structure for the TAF1 segment spanning residues 1013–1057, docked into the protein density bound to the Inr promoter element. The predicted unstructured linker regions (993–1013 and 1056–1075) are represented as dashed lines. **e**, Putative interaction between TBP and the TAND of TAF1 within the canonical state of TFIID. The low-resolution reconstruction of TFIID in the canonical state[33] is shown in mesh, superimposed on the new structure of promoter-bound TFIID. The domain organization of human TAF1 is shown at the top (the DUF3591 domain has been localized in this study).

be ~50% α-helical (Extended Data Fig. 4c–e). The residues near both termini of this missing stretch also appear to form contacts with the DNA between the Inr and MTE (Fig. 2d).

Superposition of the previous low-resolution cryo-EM reconstruction of TFIID in the canonical state[33] with our promoter-bound TFIID structure reveals that, in the former, lobe A is attached to lobe C near the newly identified position of the TAF1/TAF7 subcomplex (Fig. 2e). The N-terminal domain of TAF1 (TAND) contains two subdomains known to bind the concave and convex surfaces of TBP, respectively, thereby competing with TBP binding to DNA and TFIIA[23,37]. It is likely that in the canonical state of TFIID, TBP is at least partly inhibited from binding to promoter DNA through interactions with the TAND. In the rearranged promoter-bound state of TFIID, however, TBP is at the opposite end of the core promoter from the identified region of TAF1. Binding of TFIIA to the convex side of TBP probably plays a role in releasing TBP from inhibition by the TAND[23,37]. Additionally, TFIIA contributes to the localization of TBP in the rearranged state of TFIID through its interaction with lobe B. Thus, our studies suggest that the conformational rearrangement of TFIID and the binding of TFIIA are coupled and play critical roles in modulating the handoff of TBP to the upstream promoter region.

Previous studies suggested that TAF1 interacts with promoter DNA as a complex with TAF2[8]. The conserved N-terminal portion of TAF2 is homologous to M1-family aminopeptidases[38] and we could unambiguously assign the density adjacent to TAF1 and the downstream promoter to the TAF2 aminopeptidase-like domain (APD) by fitting the structure of the human endoplasmic reticulum aminopeptidase (ERAP1)[39] into our map (Extended Data Fig. 5a–c). We were able to generate a complete atomic model of the TAF2 APD (residues 27–975) through homology modelling and flexible fitting into the cryo-EM density (Fig. 3a, b, Extended Data Fig. 5a–d, Supplementary Video 1 and Methods). All TAF2 interactions with promoter DNA are mediated by APD domain 3, with the primary contact involving a highly conserved loop of the β-sandwich that interacts with the MTE (Fig. 3b, c and Extended Data Fig. 5e). Contacts with TAF1 are contributed by APD domains 2 and 3, involving the regions from ~467 to ~514 and from ~558 to ~561 of TAF2 (Fig. 3c, d).

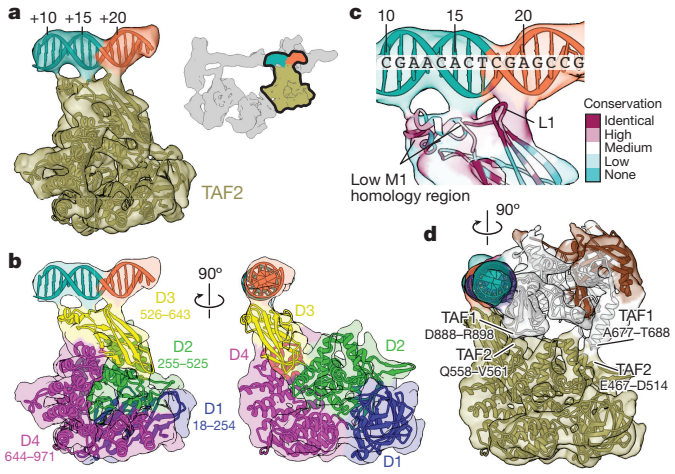


Figure 3 | TAF2 APD. a, Homology-based atomic model of the TAF2 APD fitted into the cryo-EM density. Colouring of the promoter DNA is the same as in Fig. 1. The location of the segmented density in the overall map is highlighted in the upper-right. **b**, Model of the TAF2 APD coloured by domain (D1–D4), with boundary residues for each domain indicated. **c**, Close-up of TAF2 APD domain 3 binding to promoter DNA with residues coloured according to conservation (see Extended Data Fig. 5e). **d**, Side view highlighting the TAF1–TAF2 interface, with potential regions of interaction between the two subunits indicated.

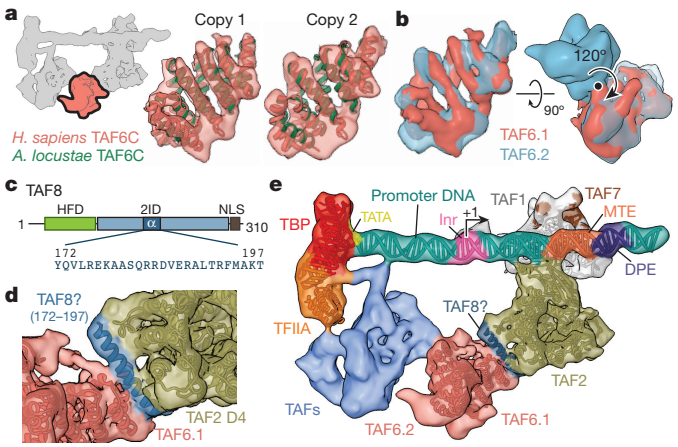


Figure 4 | Structural TAFs within lobe C. a, Docking of the crystal structure of *A. locustae* TAF6C (PDB accession number 4ATG)[22] into two adjacent densities in the cryo-EM map, termed copy 1 and copy 2. The location of the segmented density in the overall map is highlighted in the schematic on the left. **b**, The density for the two copies of TAF6C in the improved lobe C map are shown superimposed (left), and the homodimer interface and symmetry operation is depicted using the original map from Fig. 1b (right). **c**, Location and sequence of the predicted 26-residue helix within the TAF2-interacting domain (2ID) of TAF8. The relative locations of the histone fold domain (HFD) and nuclear localization signal (NLS) are also depicted. **d**, Docking of the TAF8 26 residue helix between TAF2 APD domain 4 (D4) and TAF6 copy 1 (TAF6.1). **e**, Overall architecture of TFIID with all fitted atomic models.

Structural TAFs and unassigned density

We were able to assign the majority of the remaining lobe C density to two copies of the carboxy (C)-terminal HEAT repeat domain of TAF6 (TAF6C) by fitting the crystal structure of the *Antonospora locustae* orthologue[22] (Fig. 4a, Extended Data Fig. 6a, b and Supplementary Video 1). We propose that this part of TAF6 forms a homodimer that effectively bridges the downstream promoter-interacting TAFs (TAF1, -2, and -7) with lobe B (Fig. 4b, e). The TAF6C density, at <9 Å resolution, was sufficient to unambiguously confirm the alignment with all ten α-helices in the crystal structure (Fig. 4a), and the density at the C-terminal region of both TAF6C copies is indicative of the presence of additional C-terminal α-helices, which are predicted to exist in the human protein but are missing in the crystallized orthologue (Extended Data Fig. 6b, c). We were unable to detect density near either copy of the TAF6C homodimer for the N-terminal histone fold domain of TAF6, which forms a heterodimer with the histone fold domain of TAF9 (ref. 15). This result, which suggests that the TAF6 histone fold is flexibly attached and not critical to the structural integrity of the core TFIID, agrees with the finding that the human isoform TAF6δ, which lacks a critical part of its histone fold domain, integrates into an active TFIID complex that retains all TAFs except TAF9 (ref. 40).

After accounting for the portions of TAF2 and TAF6 that we could model into lobe C, there remains clear density for two additional α-helices bridging TAF2 and TAF6 that we were not able to assign to either of these TAFs (Extended Data Fig. 6d). We propose that these helices are contributed by the C-terminal region of TAF8 because (1) TAF8 associates directly with TAF2 and mediates its nuclear import and incorporation into TFIID through its C-terminal region[25], (2) a fragment of the region critical for TAF2 binding (residues ~140-200)[25] is predicted to harbour a ~26 residue α-helix, the length of the longer helical density we observe bridging TAF2 and TAF6 (Fig. 4c–e, Extended Data Fig. 6e and Supplementary Video 1), and (3) TAF8 exhibits robust crosslinking to TAF6 within reconstituted TFIID subcomplexes[25].

We were unable to localize the positions for the rest of the TAFs (TAF3, -4, -5, -9, -10, -11, -12, and -13) within our promoter-bound

TFIID structure. These TAFs must therefore reside within lobes A2 and B, which are not yet resolved at enough resolution for reliable identification of subunits via docking of the available atomic models. The combined volume of these unassigned lobes is consistent with the ~300 kDa of structured TAF domains that have yet to be localized (Extended Data Table 1), considering that much of TFIID is predicted to be intrinsically disordered. A previously described recombinant 5TAF subcomplex contains two copies each of TAF4, -5, -6, -9, and -12 (refs 32, 41). TAF6 is the only component of 5TAF that we were able to localize in our map. While lobe B is contiguous with TAF6C, the density for lobe B is not large enough to accommodate two copies of each TAF in the 5TAF subcomplex. Additionally, we do not observe density within lobe B for the distinctive WD40 beta-propeller domain of TAF5 nor the TAF6–TAF9 histone-fold heterodimer, both of which were proposed to contact TAF6C through opposing interfaces within the recombinant 5TAF subcomplex. We therefore conclude that the components of 5TAF are probably divided between lobes B and A2 in the full TFIID complex.

TAF–less PIC structure and full PIC model

To gain structural insight into the full PIC assembly, we solved the cryo-EM structure of a simplified, TAF-less PIC containing TBP, Pol II, TFIIA, -B, -F, -S, -E, -H, and SCP DNA. This human TAF-less PIC cryo-EM reconstruction is similar to the one we previously reported using negative stain EM[42], but promoter DNA is now visible (Fig. 5a and Supplementary Video 1). By superimposing the common elements between the TFIID-IIA–SCP complex and the TAF-less PIC (that is, TBP, TFIIA, promoter DNA), we were able to generate a model of a complete TFIID-based PIC (Fig. 5b and Supplementary Video 1). Overall the two structures fit well with each other, with significant shape complementarity and minimal steric clashes, which are, however, of potential functional relevance. Superposition of the unmasked TFIID-IIA–SCP reconstruction (including lobe A2) onto our model of the TFIID-based PIC indicates that the observed range of positions for the flexible lobe A2 is overall compatible with the model, without any major clashes (Extended Data Fig. 7a). The proximity of TFIIF and TFIIE to lobe B of TFIID in our model supports the finding that these factors can interact with TAFs[11–13], thus implicating TAFs in the recruitment of PIC components (Fig. 5b, c). The unidentified density emanating from TFIIA and lobe B of TFIID and contacting the DNA downstream of the TATA sequence in the TFIID-IIA–SCP reconstruction overlaps the promoter-binding site of the RAP30 WH domain of TFIIF in the TAF-less PIC, suggesting that a structural reorganization occurs in this region upon TFIIF recruitment to the PIC (Fig. 5c).

Our model shows that Pol II docks between the up- and downstream promoter-binding regions of TFIID. The protein density bound to the Inr promoter element, which we attribute to TAF1, docks into the cleft of Pol II, adjacent to its RPB1, RPB2, and RPB5 subunits (Extended Data Fig. 7b). Downstream of the Inr, TAF1, TAF2, and the XPB subunit of TFIIH make complementary promoter contacts on opposite faces of the DNA duplex (Fig. 5b, right). Minor clashes between TAF1 and RPB1, -2, and -5 signify that this region of TFIID undergoes structural rearrangement upon loading of Pol II onto promoter DNA. Additionally, the path of the promoter DNA in the TFIID-IIA–SCP complex deviates from that seen in the TAF-less PIC, especially downstream of the TSS, further supporting a structural rearrangement in TFIID and downstream promoter DNA during PIC assembly (Extended Data Fig. 7c).

As Pol II reads through the DNA downstream of the TSS, TAF1 and TAF2 must disengage from the downstream promoter DNA before Pol II clears the promoter. Indeed, it has been found that upon recruitment of Pol II, promoter-bound TFIID undergoes an isomerization in which the TFIID contacts with promoter DNA downstream of the +10 position are released concomitantly with the engagement of the promoter DNA with Pol II upstream of this position[43].

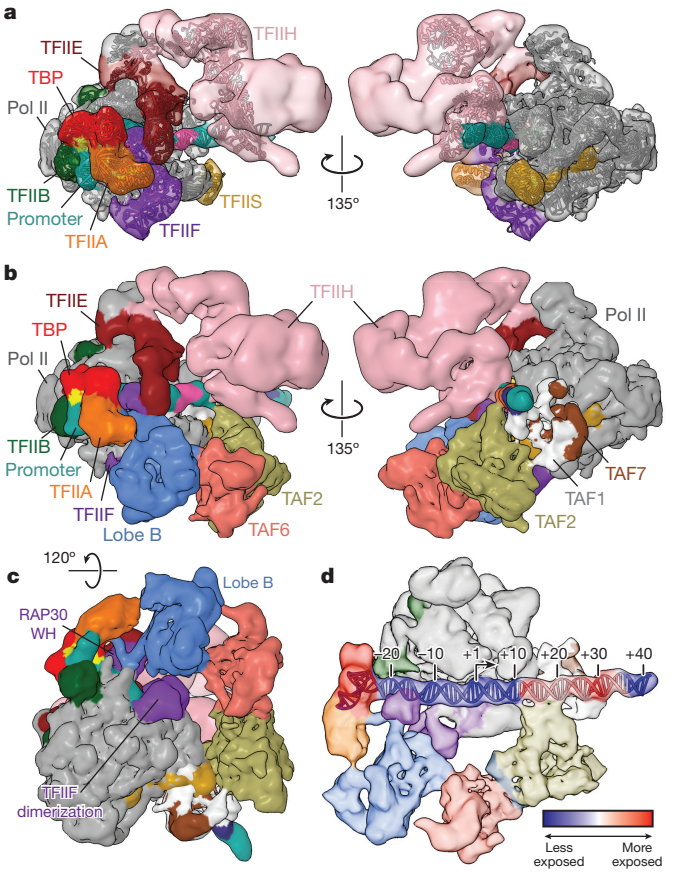


Figure 5 | Model of the TFIID-based PIC. a, Cryo-EM reconstruction of the human TAF-less PIC, with fitted atomic models. Views are similar to those in Fig. 5a in ref. 42. **b**, Model of the TFIID-based PIC generated by superimposing the densities for TBP, TFIIA, and promoter DNA within the TFIID-IIA–SCP and TAF-less PIC reconstructions. For clarity, the superimposed densities from the TAF-less PIC reconstruction are hidden. **c**, Bottom view of the TFIID-based PIC model highlighting putative interactions between lobe B of TFIID and TFIIF. **d**, Changes in protein–DNA contacts following the addition of TFIIB–Pol II–TFIIF to the TFIID-IIA–SCP complex, according to data published in ref. 43. The blue to red colouring scale represents the rate constant of change in DNaseI cleavage (k_{obs}) for each base pair following the addition of Pol II–TFIIB–TFIIF, with blue set to $-10 \times 10^{-3}\,s^{-1}$, corresponding to regions that become more protected, and red set to $+10 \times 10^{-3}\,s^{-1}$, corresponding to regions that become more exposed.

Those findings can be mapped onto our model (Fig. 5d) and agree with our structure-based proposal of a reorganization in the downstream region of the PIC. It has also been recently demonstrated that a chemical inhibitor of this isomerization interacts with the intrinsically disordered region of TAF2 and prevents the first round of transcription initiation by blocking the initial recruitment of Pol II[44]. However, the inhibitor has no effect on reinitiation of transcription, suggesting that the isomerization does not occur during reloading of Pol II. We propose that the isomerization of promoter-bound TFIID required for Pol II recruitment, its engagement with promoter DNA, and its clearance of the promoter during transcription initiation, largely involves the release of downstream promoter contacts by TAF1 and TAF2. Since the same isomerization does not take place during reinitiation, it is likely that TAF1 and TAF2 do not re-form some of these promoter contacts following the first round of initiation. The release of TAF7 from TFIID following PIC assembly has been shown to be required for transcription initiation[45], and could potentially serve as the mechanism for preventing re-engagement of the promoter DNA by TAF1 and TAF2.

Recent cryo-EM studies have revealed the binding site of the yeast Mediator complex on Pol II within a minimal transcription initiation complex[46,47]. Superposition of this structure with our TFIID-based PIC model shows that TFIID and Mediator occupy opposite faces of Pol II (Extended Data Fig. 7d, e and Supplementary Video 1).

Role of TFIID in transcription initiation

Our structures suggest that a primary function of TFIID during PIC assembly is the proper positioning of TBP on the upstream promoter, which ultimately determines the placement of Pol II relative to the TSS. For the majority of human promoters, which lack a canonical TATA sequence[48–50], accurate loading of TBP would be ensured by the TAF subunits of TFIID, which collectively act as a molecular ruler to position TBP at a location on the upstream promoter that is precisely defined by the downstream promoter-binding sites of TAF1–TAF2 and the length of the BC core of TFIID. TAFs may also facilitate PIC assembly by contributing to the incorporation of TFIIF and TFIIE to the growing PIC. To accommodate the recruitment of Pol II and its subsequent engagement with promoter DNA, an isomerization occurs in which TAF1 and TAF2 probably release some of their contacts with the downstream DNA. Our model suggests that TAFs are generally required for initial PIC assembly and first round of transcription initiation, but that at least some TAFs may be dispensable for the reloading of Pol II.

TAFs are probably also critical for providing additional levels of control of transcription initiation. For instance, competition from the TAND of TAF1 with the binding of TFIIA and DNA to TBP, which we propose involves the conformational rearrangement of TFIID, may present additional opportunities for regulating PIC assembly. The timing and rate of PIC assembly at the promoter will ultimately be regulated through combinatorial interactions involving TAFs, variable promoter sequences, activators, and epigenetic marks. The demonstrated flexible character of TFIID is likely to be an important property for integrating regulatory cues and allowing sequential conformational states that provide checkpoints throughout the processes of PIC assembly and transcription initiation. The structures presented here offer a structural framework for understanding the complex mechanism underlying TFIID function, shedding new light into the overlapping roles of TFIID as both a coactivator and a general platform for PIC assembly in the coordination of transcription initiation.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 24 October 2015; accepted 3 February 2016.
Published online 23 March 2016.

1. Levine, M., Cattoglio, C. & Tjian, R. Looping back to leap forward: transcription enters a new era. *Cell* **157,** 13–25 (2014).
2. Thomas, M. C. & Chiang, C. M. The general transcription machinery and general cofactors. *Crit. Rev. Biochem. Mol. Biol.* **41,** 105–178 (2006).
3. Buratowski, S., Hahn, S., Guarente, L. & Sharp, P. A. Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell* **56,** 549–561 (1989).
4. Burley, S. K. & Roeder, R. G. Biochemistry and structural biology of transcription factor IID (TFIID). *Annu. Rev. Biochem.* **65,** 769–799 (1996).
5. Albright, S. R. & Tjian, R. TAFs revisited: more data reveal new twists and confirm old ideas. *Gene* **242,** 1–13 (2000).
6. Verrijzer, C. P., Chen, J. L., Yokomori, K. & Tjian, R. Binding of TAFs to core elements directs promoter selectivity by RNA polymerase II. *Cell* **81,** 1115–1125 (1995).
7. Burke, T. W. & Kadonaga, J. T. The downstream core promoter element, DPE, is conserved from *Drosophila* to humans and is recognized by TAFII60 of *Drosophila*. *Genes Dev.* **11,** 3020–3031 (1997).
8. Chalkley, G. E. & Verrijzer, C. P. DNA binding site selection by RNA polymerase II TAFs: a TAF(II)250-TAF(II)150 complex recognizes the initiator. *EMBO J.* **18,** 4835–4845 (1999).
9. Lee, D. H. *et al.* Functional characterization of core promoter elements: the downstream core element is recognized by TAF1. *Mol. Cell. Biol.* **25,** 9674–9686 (2005).
10. Theisen, J. W., Lim, C. Y. & Kadonaga, J. T. Three key subregions contribute to the function of the downstream RNA polymerase II core promoter. *Mol. Cell. Biol.* **30,** 3471–3479 (2010).
11. Hisatake, K. *et al.* Evolutionary conservation of human TATA-binding-polypeptide-associated factors TAFII31 and TAFII80 and interactions of TAFII80 with other TAFs and with general transcription factors. *Proc. Natl Acad. Sci. USA* **92,** 8195–8199 (1995).
12. Ruppert, S. & Tjian, R. Human TAFII250 interacts with RAP74: implications for RNA polymerase II initiation. *Genes Dev.* **9,** 2747–2755 (1995).
13. Dubrovskaya, V. *et al.* Distinct domains of hTAFII100 are required for functional interaction with transcription factor TFIIFβ (RAP30) and incorporation into the TFIID complex. *EMBO J.* **15,** 3702–3712 (1996).
14. Wu, S. Y. & Chiang, C. M. TATA-binding protein-associated factors enhance the recruitment of RNA polymerase II by transcriptional activators. *J. Biol. Chem.* **276,** 34235–34243 (2001).
15. Xie, X. *et al.* Structural similarity between TAFs and the heterotetrameric core of the histone octamer. *Nature* **380,** 316–322 (1996).
16. Birck, C. *et al.* Human TAF(II)28 and TAF(II)18 interact through a histone fold encoded by atypical evolutionary conserved motifs also found in the SPT3 family. *Cell* **94,** 239–249 (1998).
17. Jacobson, R. H., Ladurner, A. G., King, D. S. & Tjian, R. Structure and function of a human TAFII250 double bromodomain module. *Science* **288,** 1422–1425 (2000).
18. Werten, S. *et al.* Crystal structure of a subcomplex of human transcription factor TFIID formed by TATA binding protein-associated factors hTAF4 (hTAF(II)135) and hTAF12 (hTAF(II)20). *J. Biol. Chem.* **277,** 45502–45509 (2002).
19. Bleichenbacher, M., Tan, S. & Richmond, T. J. Novel interactions between the components of human and yeast TFIIA/TBP/DNA complexes. *J. Mol. Biol.* **332,** 783–793 (2003).
20. Bhattacharya, S., Takada, S. & Jacobson, R. H. Structural analysis and dimerization potential of the human TAF5 subunit of TFIID. *Proc. Natl Acad. Sci. USA* **104,** 1189–1194 (2007).
21. Wang, X. *et al.* Conserved region I of human coactivator TAF4 binds to a short hydrophobic motif present in transcriptional regulators. *Proc. Natl Acad. Sci. USA* **104,** 7839–7844 (2007).
22. Scheer, E., Delbac, F., Tora, L., Moras, D. & Romier, C. TFIID TAF6-TAF9 complex formation involves the HEAT repeat-containing C-terminal domain of TAF6 and is modulated by TAF5 protein. *J. Biol. Chem.* **287,** 27580–27592 (2012).
23. Anandapadamanaban, M. *et al.* High-resolution structure of TBP with TAF1 reveals anchoring patterns in transcriptional regulation. *Nature Struct. Mol. Biol.* **20,** 1008–1014 (2013).
24. Wang, H., Curran, E. C., Hinds, T. R., Wang, E. H. & Zheng, N. Crystal structure of a TAF1-TAF7 complex in human transcription factor IID reveals a promoter binding module. *Cell Res.* **24,** 1433–1444 (2014).
25. Trowitzsch, S. *et al.* Cytoplasmic TAF2–TAF8–TAF10 complex provides evidence for nuclear holo-TFIID assembly from preformed submodules. *Nature Commun.* **6,** 6011 (2015).
26. Andel, F., III, Ladurner, A. G., Inouye, C., Tjian, R. & Nogales, E. Three-dimensional structure of the human TFIID-IIA-IIB complex. *Science* **286,** 2153–2156 (1999).
27. Brand, M., Leurent, C., Mallouh, V., Tora, L. & Schultz, P. Three-dimensional structures of the TAFII-containing complexes TFIID and TFTC. *Science* **286,** 2151–2153 (1999).
28. Leurent, C. *et al.* Mapping histone fold TAFs within yeast TFIID. *EMBO J.* **21,** 3424–3433 (2002).
29. Grob, P. *et al.* Cryo-electron microscopy studies of human TFIID: conformational breathing in the integration of gene regulatory cues. *Structure* **14,** 511–520 (2006).
30. Liu, W. L. *et al.* Structural changes in TAF4b-TFIID correlate with promoter selectivity. *Mol. Cell* **29,** 81–91 (2008).
31. Papai, G. *et al.* TFIIA and the transactivator Rap1 cooperate to commit TFIID for transcription initiation. *Nature* **465,** 956–960 (2010).
32. Bieniossek, C. *et al.* The architecture of human general transcription factor TFIID core complex. *Nature* **493,** 699–702 (2013).
33. Cianfrocco, M. A. *et al.* Human TFIID binds to core promoter DNA in a reorganized structural state. *Cell* **152,** 120–131 (2013).
34. Juven-Gershon, T., Cheng, S. & Kadonaga, J. T. Rational design of a super core promoter that enhances gene expression. *Nature Methods* **3,** 917–922 (2006).
35. Kraemer, S. M., Ranallo, R. T., Ogg, R. C. & Stargell, L. A. TFIIA interacts with TFIID via association with TATA-binding protein and TAF40. *Mol. Cell. Biol.* **21,** 1737–1746 (2001).
36. Gajiwala, K. S. & Burley, S. K. Winged helix proteins. *Curr. Opin. Struct. Biol.* **10,** 110–116 (2000).
37. Kokubo, T., Swanson, M. J., Nishikawa, J. I., Hinnebusch, A. G. & Nakatani, Y. The yeast TAF145 inhibitory domain and TFIIA competitively bind to TATA-binding protein. *Mol. Cell. Biol.* **18,** 1003–1012 (1998).
38. Malkowska, M., Kokoszynska, K., Rychlewski, L. & Wyrwicz, L. Structural bioinformatics of the general transcription factor TFIID. *Biochimie* **95,** 680–691 (2013).
39. Kochan, G. *et al.* Crystal structures of the endoplasmic reticulum aminopeptidase-1 (ERAP1) reveal the molecular basis for N-terminal peptide trimming. *Proc. Natl Acad. Sci. USA* **108,** 7745–7750 (2011).

40. Bell, B., Scheer, E. & Tora, L. Identification of hTAF(II)80 delta links apoptotic signaling pathways to transcription factor TFIID function. *Mol. Cell* **8,** 591–600 (2001).
41. Wright, K. J., Marr, M. T., II & Tjian, R. TAF4 nucleates a core subcomplex of TFIID and mediates activated transcription from a TATA-less promoter. *Proc. Natl Acad. Sci. USA* **103,** 12347–12352 (2006).
42. He, Y., Fang, J., Taatjes, D. J. & Nogales, E. Structural visualization of key steps in human transcription initiation. *Nature* **495,** 481–486 (2013).
43. Yakovchuk, P., Gilman, B., Goodrich, J. A. & Kugel, J. F. RNA polymerase II and TAFs undergo a slow isomerization after the polymerase is recruited to promoter-bound TFIID. *J. Mol. Biol.* **397,** 57–68 (2010).
44. Zhang, Z. *et al.* Chemical perturbation of an intrinsically disordered region of TFIID distinguishes two modes of transcription initiation. *eLife* **4,** e07777 (2015).
45. Gegonne, A., Devaiah, B. N. & Singer, D. S. TAF7: traffic controller in transcription initiation. *Transcription* **4,** 29–33 (2013).
46. Plaschka, C. *et al.* Architecture of the RNA polymerase II-Mediator core initiation complex. *Nature* **518,** 376–380 (2015).
47. Tsai, K. L. *et al.* Subunit architecture and functional modular rearrangements of the transcriptional mediator complex. *Cell* **157,** 1430–1444 (2014).
48. Kim, T. H. *et al.* A high-resolution map of active promoters in the human genome. *Nature* **436,** 876–880 (2005).
49. Carninci, P. *et al.* Genome-wide analysis of mammalian promoter architecture and evolution. *Nature Genet.* **38,** 626–635 (2006).
50. Cooper, S. J., Trinklein, N. D., Anton, E. D., Nguyen, L. & Myers, R. M. Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome. *Genome Res.* **16,** 1–10 (2006).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank P. Grob, S. Howes, and R. Zhang for electron microscopy support; T. Houweling for computer support; S. Scheres for technical advice on image processing; A. Patel for discussion; C. Inouye for providing us with recombinant TFIIF; S. Zheng for providing TAF4 mAb; and D. King for providing peptides. We are thankful to R. Tjian and J. Kadonaga for their comments on the manuscript. Computational resources were provided in part by the National Energy Research Scientific Computing Center (DE-AC02-05CH11231). This work was funded by NIGMS (GM63072 to E.N.) and Spanish Ministry of Economy and Competitiveness (BFU2013-44306P to P.C.). R.K.L. was supported by the NIGMS Molecular Biophysics Training Grant (GM008295). E.N. is a Howard Hughes Medical Institute Investigator.

Author Contributions R.K.L. and Y.H. designed and performed the experiments; J.R.L.-B. and P.C. performed structural modelling; J.F. purified the TFIID, Pol II, TFIIE, and TFIIH; R.L., Y.H., and E.N. analysed the data and wrote the paper.

Author Information Cryo-EM density maps have been deposited in the Electron Microscopy Data Bank (EMDB) under codes EMD-3304 (TFIID-IIA–SCP complex), EMD-3305 (locally refined BC-core of TFIID-IIA–SCP complex), EMD-3306 (locally refined lobe C of TFIID-IIA–SCP complex), and EMD-3307 (TAF-less PIC). Model coordinates have been deposited in the Protein Data Bank (PDB) under accession number 5FUR. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.N. (ENogales@lbl.gov).

METHODS

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

Preparation of transcription complexes for cryo-EM. TFIID, Pol II, and TFIIH were immunopurified from HeLa cell nuclear extracts following previously established protocols[51,52]. The human TFIIA used in the reconstitution of both the TFIID-IIA–SCP complex and TAF-less PIC was recombinantly expressed and purified first as three separate polypeptides (TFIIAα 2–58, TFIIAβ 303–376, and TFIIAγ 2–109) from *Escherichia coli*, then reconstituted into the conserved three-subunit core TFIIA similarly as in ref. 19. The C-terminal DNA-binding core of human TBP (residues 159–339), and full-length TFIIB, TFIIE, TFIIF, and TFIIS, were used in the reconstitution of the TAF-less PIC and were recombinantly expressed and purified from *E. coli*.

The sequence of the promoter DNA construct used in this study was based on the SCP[34], except that a BREu element was introduced upstream of the TATA box and an EcoRI restriction enzyme site was included downstream of the DPE element for purification purposes (template 5′-ACTGGGGAATTCCATGGTCCGTAGG CACGTCTGCTCGGCTCGAGTGTTCGATCGCGACTGAGGACGAACGCG CCCCCACCCCCTTTTATAGGCGCCCTTC-3′; non-template 5′-GAAGGGC GCCTATAAAAGGGGGTGGGGGCGCGTTCGTCCTCAGTCGCGATCGAA CACTCGAGCCGAGCAGACGTGCCTACGGACCATGGAATTCCCCAGT-3′). The template strand was modified with a biotin tag at the 5′ end (Integrated DNA Technologies). To generate the duplexed DNA, equimolar amounts of the template and non-template strand oligonucleotides were mixed at a final concentration of 25 μM in water, and the annealing reaction was heated to 95 °C for 5 min then allowed to slowly cool to room temperature over 2–3 h.

To assemble TFIID-IIA–SCP complex, 5.0 pmol TFIIA was first added to 2.5 pmol purified HeLa TFIID in assembly buffer (20 mM HEPES pH 7.9, 0.2 mM EDTA, 10% glycerol, 6 mM $MgCl_2$, 80 mM KCl, 1 mM DTT, 0.05% NP-40) and incubated for 5 min at 37 °C. A limiting amount of biotinylated SCP DNA (1 pmol) was then added and the assembly reaction was incubated for 10 min at 37 °C. The reaction was added to 0.25 μl Streptavidin Mag Sepharose magnetic beads (GE Healthcare) and incubated at 28 °C for 15 min. The beads were then washed three times with washing buffer (10 mM HEPES pH 7.9, 3% trehalose, 8 mM $MgCl_2$, 100 mM KCl, 1 mM DTT, 0.025% NP-40). The promoter-bound complex was then eluted by incubating in 3.6 μl of elution buffer (10 mM HEPES pH 7.9, 10 mM $MgCl_2$, 3% trehalose, 50 mM KCl, 1 mM DTT, 0.05% NP-40, 1 unit per microlitre EcoRI-HF (New England BioLabs)) for 30 min at 37 °C.

The TAF-less PIC was assembled similarly as before[42], except for an additional incubation of TFIIS at a final concentration of 200 nM with the purified PIC before application to the grid. We included TFIIS in our initiation assemblies because of its novel role in active PIC formation besides that in elongation[53].

Following the restriction digest elution, purified TFIID-IIA–SCP complex or TAF-less PIC was crosslinked with 0.01% glutaraldehyde for 5 min on ice, then used immediately for cryo-EM sample preparation.

Electron microscopy. Cryo-EM samples were prepared on continuous carbon coated C-flat holey carbon grids (Protochips). Grids were plasma cleaned for 10 s in air using a Solarus Plasma Cleaner (Gatan) operating at 10 W. Immediately following crosslinking, 4 μl of purified TFIID-IIA–SCP complex or TAF-less PIC was added to the plasma-cleaned grid and loaded into a Vitrobot (FEI). The sample was incubated on the grid for 10 min at 4 °C and 100% relative humidity to enhance its absorption onto the carbon substrate, then was blotted and immediately plunge-frozen in liquid ethane. Frozen grids were transferred to a 626 Cryo-Transfer Holder (Gatan) and loaded into a Titan electron microscope (FEI) operating at 300 keV. Images were recorded with a K2 direct electron detector (Gatan) operating in counting mode at a calibrated magnification of 37,879 (1.32 Å per pixel) and a defocus range of −2 μm to −4 μm, using the Leginon data collection software for semi-automated acquisition targeting. Twenty-frame exposures were taken at 0.5 s per frame (10 s total exposure time), using a dose rate of 8 electrons per pixel per second (4.6 electrons per square ångström per second or 2.3 electrons per square ångström per frame), corresponding to a total dose of 46 electrons per square ångström per micrograph.

Image processing. The exposure frames were aligned using MotionCorr[54] to correct for specimen motion, and the average of the aligned frames was used for initial processing. The CTF parameters of the micrographs were estimated using CTFFIND3[55]. For the TFIID-IIA–SCP complex, RELION[56] (version 1.4-beta) was used for automatic selection of 203,163 particles from 1,253 micrographs (Extended Data Fig. 1a). For the TAF-less PIC, 245,501 particles were automatically selected from 855 micrographs using a difference of Gaussians (DoG) particle picker[57] within the Appion image processing environment[58]. All two- and three-dimensional classification and refinement steps were performed within RELION[56] (version 1.4-beta).

For the TFIID-IIA–SCP complex, the initial set of 203,163 particles was subjected to an initial three-dimensional classification, with the negative stain reconstruction of the same complex (in which the nucleic acid is not visible) low-pass filtered to 60 Å used as an initial reference (Extended Data Fig. 1b). Three out of five classes in this classification, corresponding to 121,459 particle images, were indicative of promoter-bound complexes and were selected for further processing. Reference-free two-dimensional classification of this set of images was used to select for 56,457 high-quality particle images. The 56,457-particle set was then subjected to three-dimensional refinement and particle polishing procedure within RELION[59] to correct for individual particle motion and beam-induced radiation damage of the sample. Owing to the low contrast inherent in the images of this sample, the per-frame *B*-factor plot used to model the beam-induced radiation damage was too noisy to use for modelling. Instead, we generated an idealized curve for the dose-dependent *B*-factor on the basis of cryo-EM data collected on microtubules under similar imaging conditions (Extended Data Fig. 1c), which we applied to our data during the particle polishing step. The resulting set of 56,457 'polished' particles (56k set) was used in all subsequent three-dimensional classification and refinement.

An initial three-dimensional refinement of the 56k set resulted in a reconstruction of the promoter-bound TFIID-IIA complex at 10.2 Å resolution (Extended Data Fig. 1d, e). All resolutions reported herein correspond to the gold-standard Fourier shell correlation (FSC) = 0.143 criterion[60]. Local resolution estimation indicated that the density for lobe A2 of TFIID was at much lower resolution than the promoter-bound BC-core (Extended Data Fig. 1e). To improve the reconstruction of the promoter-bound BC-core of TFIID, the orientations of the particle images were locally refined against a reference in which a mask was applied around the BC-core density, effectively excluding the contribution of lobe A2 signal from the alignment. The images were then three-dimensionally classified within the same mask, and one class with 22,050 particles exhibited the lowest error in angular and translational alignment (Extended Data Fig. 2a). This set of 22,050 particle images was then subjected to three-dimensional refinement without using any mask, followed by a local refinement against a reference with a mask around the BC-core. This procedure resulted in an improved reconstruction of the BC-core with an overall resolution of 8.7 Å (Extended Data Fig. 2c).

Three-dimensional classification and local-resolution analysis of the BC-core density indicated further conformational heterogeneity, which could be largely characterized as mobility of lobe B, TBP/TFIIA module, and upstream promoter DNA relative to lobe C and bound downstream promoter DNA (Extended Data Fig. 2d–f). Therefore, we employed a similar strategy to improve the reconstruction of lobe C and bound downstream promoter DNA (Extended Data Fig. 2b). The resulting lobe C density incorporated 28,448 particle images and had an overall resolution of 8.2 Å (Extended Data Fig. 2c).

For the TAF-less PIC, all particle picks were used for an initial three-dimensional classification, using the previously published negative stain reconstruction of the TFIIH-containing PIC (EMDB code 2308), low-pass filtered to 60 Å, as an initial reference. A single class corresponded to the fully assembled, TFIIH-containing TAF-less PIC, comprising 24,290 particle images. A three-dimensional refinement of this set of particles yielded the final reconstruction at 7.2 Å resolution.

Local resolution estimations were performed using the Bsoft software package[61], and all final volumes shown in this paper have been automatically sharpened using the post-processing program within RELION and then filtered according to local resolution using the blocfilt program within Bsoft.

Structural modelling. For the TFIID-IIA–SCP complex, densities were initially assigned to specific components by rigid-body docking of known crystal structures (TBP, TFIIA, TATA DNA, TAF1, TAF7) or homologous structures (TAF2 and TAF6) using UCSF Chimera[62], ADP_EM[63], or Situs[64]. These structures were used as starting point for flexible refinement using iMODFIT[65] when necessary. Reliable homology models were generated with either the SWISS-MODEL server[66] or I-TASSER server[67].

The TBP–TFIIA–TATA DNA complex was the first structure to be fitted into the density, which was accomplished by docking the crystal structure of this complex[19] as a single rigid body. The model of the entire SCP was then generated by extrapolating B-form DNA from TBP-bound TATA box sequence present in the fitted crystal structure, followed by manual bending of the DNA structure using the 3D-DART server[68], and finally by flexible fitting of the DNA into the cryo-EM density using iMODFIT[65].

The density for the TAF1–TAF7 promoter-binding module was initially identified within the promoter-bound BC-core of TFIID using the unbiased 6D global docking search algorithm implemented in Situs, and the rigid body

docking of the crystal structure (PDB accession number 4RGW)[24] was further refined using the higher-resolution lobe C density within Chimera. The secondary and three-dimensional structure prediction for TAF1 993–1074 (putative Inr-binding domain) was performed using the I-TASSER server[67].

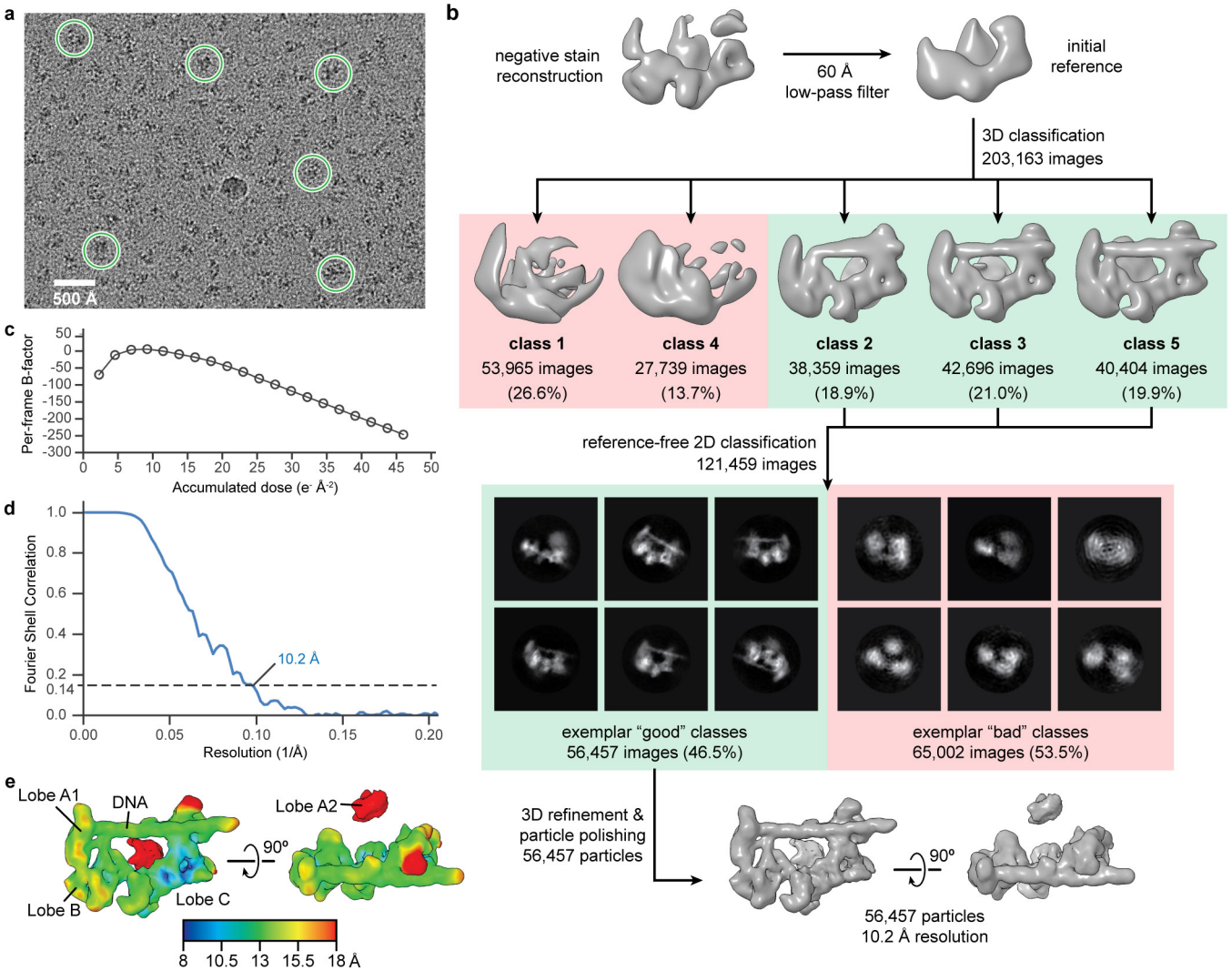
The density for TAF2 was initially identified through manual docking of the crystal structure of the human endoplasmic reticulum aminopeptidase I (ERAPI, PDB accession number 2YD0)[39] within UCSF Chimera. Homology modelling of the TAF2 APD began with the building of APD domains 1 and 2 within the SWISS-MODEL server[66], using the structures of leukotriene A-4 hydrolase (PDB accession number 3U9W: 21% sequence identity, 31% similarity, and 79% coverage) and endoplasmic reticulum aminopeptidase 2 (PDB accession number 3SE6: 16% identity, 30% similarity, and 88% coverage) as templates. While the first 26 N-terminal residues and an insertion (residues 88–124) of APD domain 1 could not be modelled by homology, there are two nearby unassigned densities in contact with APD domain 4 that are likely to account this missing modelling part. TAF2 APD domain 3 was then modelled using templates derived from aminopeptidase N (PDB accession number 3B34 and 4QME). Despite the high confidence secondary structure prediction score of the 16 predicted alpha helices in APD domain 4, the low sequence identities (between 10% and 19%), coverages (between 50% and 86%), and the conformational variability of the model templates precluded that a single model fit well in the armadillo fold visible in the density. However, we were able to accommodate the distinctive armadillo curvature by merging and flexibly fitting fragments from two to six helices extracted from different models. The templates of these models were the structures of aminopeptidase N (PDB accession number 3B34 and 4QME), deoxyhypusine hydroxylase (PDB accession number 4D4Z), AP2 clathrin adaptor (PDB accession number 1GW5), and hypothetical protein yibA (PDB accession number 1OYZ). The missing loops between fragments and domains were modelled *ab initio* using RCD[69] to obtain a complete model. Finally, the full TAF2 APD model was relaxed with Chiron[70] and PyRosetta[71] to prevent clashes and to improve geometry.

The DUF1546 domain of TAF6 (TAF6C) was modelled from the crystal structure of the *A. locustae* orthologue (PDB accession number 4ATG: 23% identity and 81% coverage)[22]. Two copies of this well conserved structure containing HEAT repeats were easily located by rigid-body docking in the cryo-EM density using Situs[64]. In the contact region between TAF6 and the TAF2 APD domain 4, there is obvious density for two alpha helices, which we predict to be derived from TAF8 on the basis of biochemical data and secondary structure prediction.

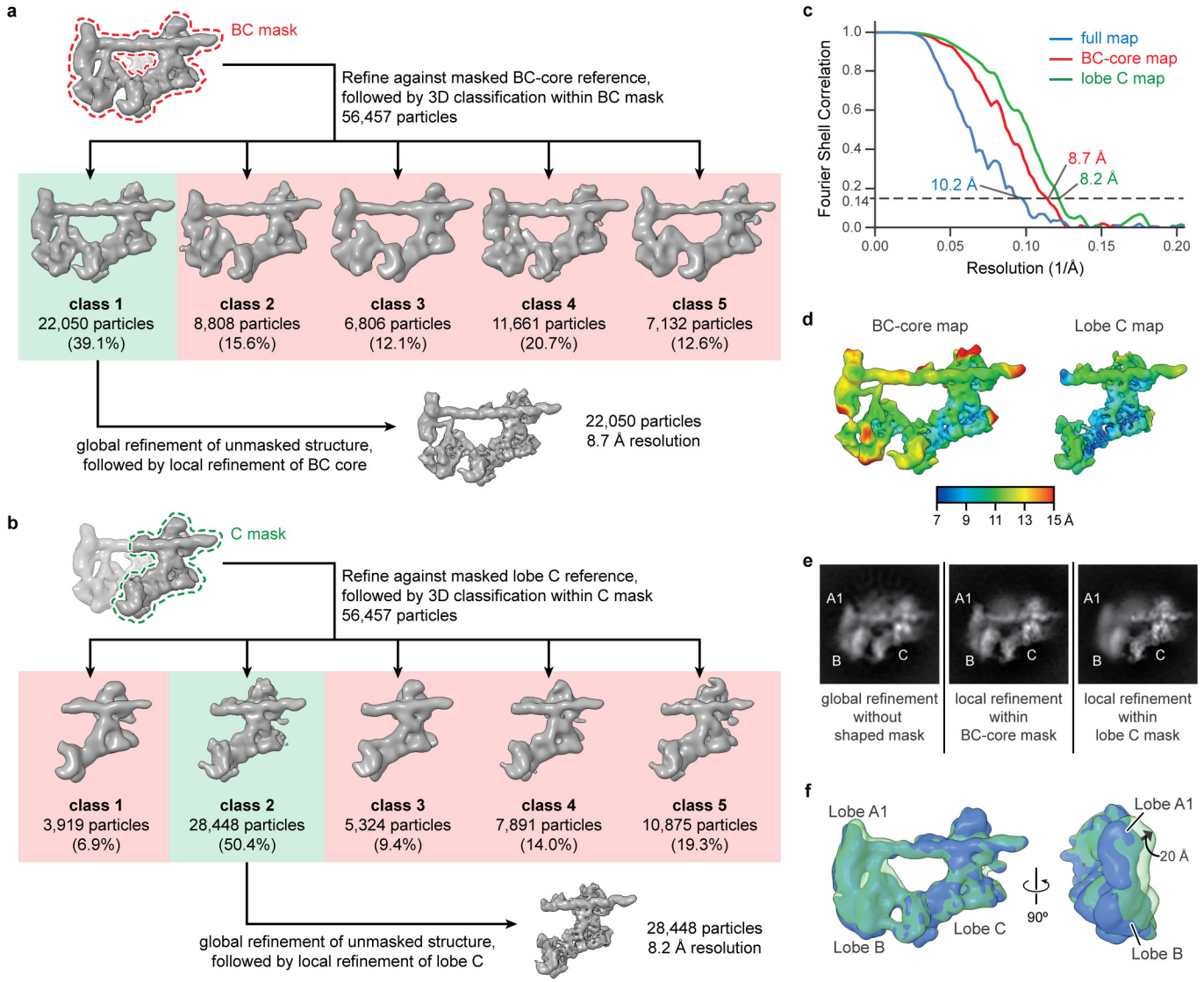
The model for the TAF-less PIC is based on the previously published model using the negative stain EM reconstruction of the PIC containing TBP, TFIIA, SCP DNA, TFIIB, Pol II, TFIIF, TFIIE, and TFIIH. The model for TFIIS was based on a combination of the yeast Pol II–TFIIS complex (PDB accession number 1Y1V) and the free human TFIIS domain II (PDB accession number 3NDQ). The model of the TFIID-based PIC was generated by superimposing the promoter DNA downstream of the TATA sequence in the TFIID-IIA–SCP structure with that in the TAF-less PIC structure within UCSF Chimera.

All multi-sequence alignments were performed using Clustal Omega[72]. Secondary structure predictions were performed with the PSIPRED server[73], except where noted. All molecular graphics and analyses were performed with the UCSF Chimera package[62].

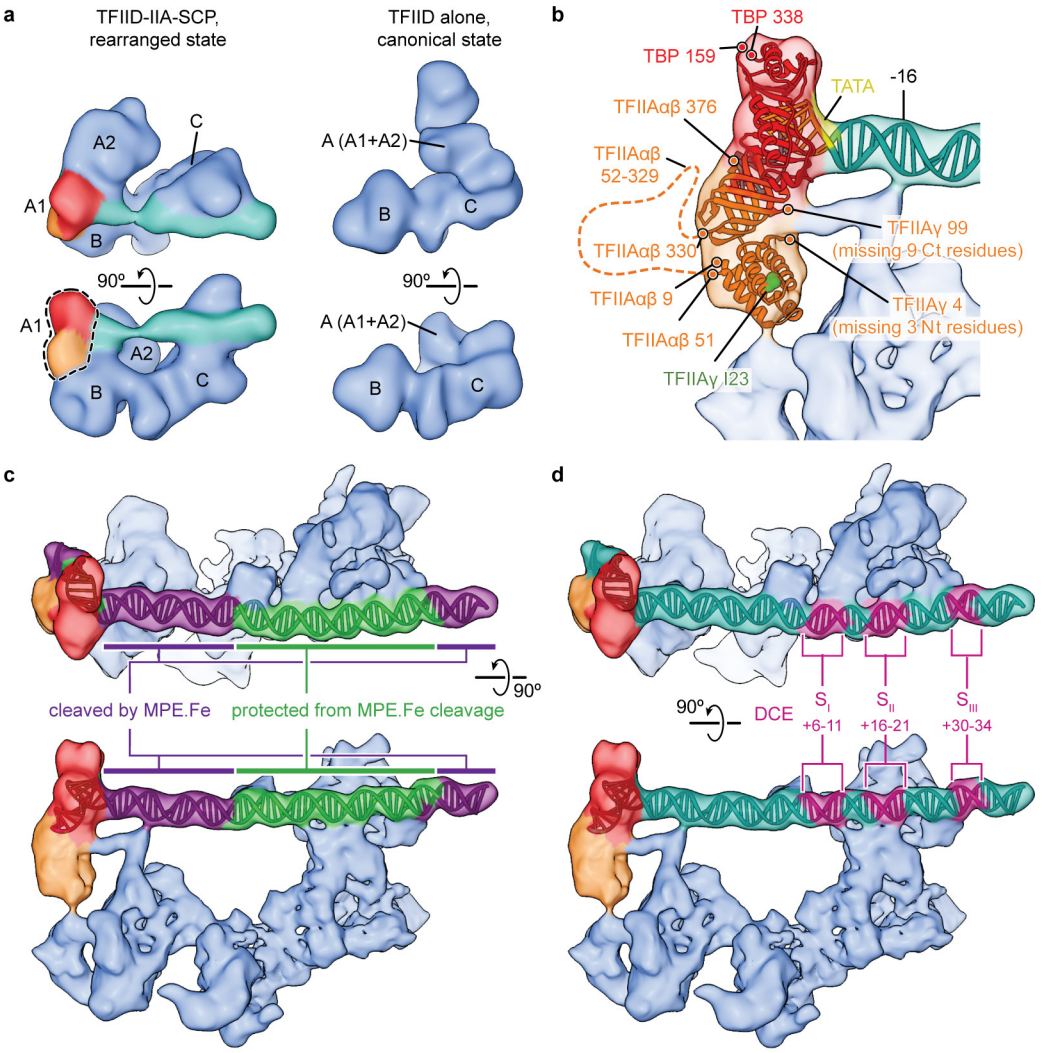
51. Pal, M., Ponticelli, A. S. & Luse, D. S. The role of the transcription bubble and TFIIB in promoter clearance by RNA polymerase II. *Mol. Cell* **19,** 101–110 (2005).
52. Revyakin, A. *et al.* Transcription initiation by human RNA polymerase II visualized at single-molecule resolution. *Genes Dev.* **26,** 1691–1702 (2012).
53. Kim, B. *et al.* The transcription elongation factor TFIIS is a component of RNA polymerase II preinitiation complexes. *Proc. Natl Acad. Sci. USA* **104,** 16068–16073 (2007).
54. Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature Methods* **10,** 584–590 (2013).
55. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142,** 334–347 (2003).
56. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180,** 519–530 (2012).
57. 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).
58. Lander, G. C. *et al.* Appion: an integrated, database-driven pipeline to facilitate EM image processing. *J. Struct. Biol.* **166,** 95–102 (2009).
59. Scheres, S. H. Beam-induced motion correction for sub-megadalton cryo-EM particles. *eLife* **3,** e03665 (2014).
60. Henderson, R. *et al.* Outcome of the first electron microscopy validation task force meeting. *Structure* **20,** 205–214 (2012).
61. Heymann, J. B. Bsoft: image and molecular processing in electron microscopy. *J. Struct. Biol.* **133,** 156–169 (2001).
62. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).
63. Garzón, J. I., Kovacs, J., Abagyan, R. & Chacón, P. ADP_EM: fast exhaustive multi-resolution docking for high-throughput coverage. *Bioinformatics* **23,** 427–433 (2007).
64. Wriggers, W. Using Situs for the integration of multi-resolution structures. *Biophys. Rev.* **2,** 21–27 (2010).
65. Lopéz-Blanco, J. R. & Chacón, P. iMODFIT: efficient and robust flexible fitting based on vibrational analysis in internal coordinates. *J. Struct. Biol.* **184,** 261–270 (2013).
66. Biasini, M. *et al.* SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **42,** W252–W258 (2014).
67. Yang, J. *et al.* The I-TASSER Suite: protein structure and function prediction. *Nature Methods* **12,** 7–8 (2015).
68. van Dijk, M. & Bonvin, A. M. 3D-DART: a DNA structure modelling server. *Nucleic Acids Res.* **37,** W235–W239 (2009).
69. Chys, P. & Chacón, P. Random coordinate descent with spinor-matrices and geometric filters for efficient loop closure. *J. Chem. Theory Comput.* **9,** 1821–1829 (2013).
70. Ramachandran, S., Kota, P., Ding, F. & Dokholyan, N. V. Automated minimization of steric clashes in protein structures. *Proteins* **79,** 261–270 (2011).
71. Chaudhury, S., Lyskov, S. & Gray, J. J. PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics* **26,** 689–691 (2010).
72. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7,** 539 (2011).
73. Buchan, D. W., Minneci, F., Nugent, T. C., Bryson, K. & Jones, D. T. Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res.* **41,** W349–W357 (2013).



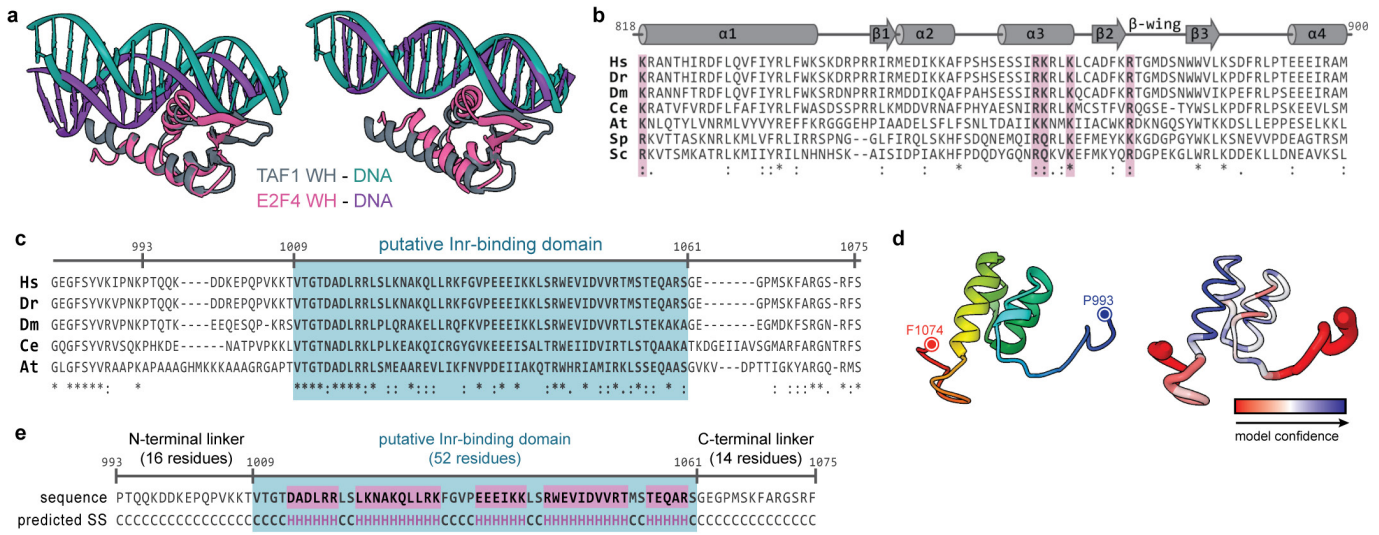
Extended Data Figure 1 | Cryo-EM of the TFIID-IIA–SCP complex. **a**, Representative micrograph of frozen-hydrated TFIID-IIA–SCP complexes. Examples of particle picks are indicated by the green circles; 203,163 such picks were made from 1,253 total micrographs. **b**, Initial classification and refinement scheme for the TFIID-IIA–SCP structure (see Methods). **c**, Idealized dose-dependent *B*-factor plot based on cryo-EM data collected on microtubules under similar imaging conditions. This plot was used for the particle polishing step in **b**. **d**, **e**, Fourier shell correlation plot (**d**) and local resolution estimation (**e**) for the final reconstruction shown in **b**.



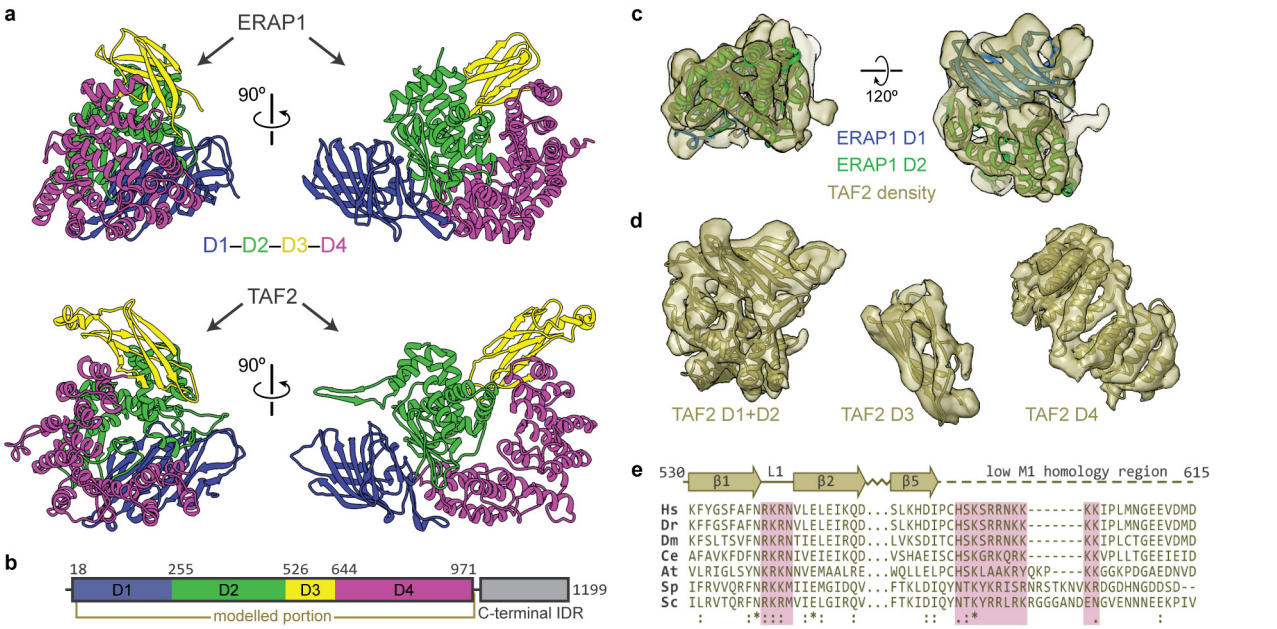
Extended Data Figure 2 | Focused classification and refinement of the promoter-bound BC-core and lobe C of TFIID. a, **b**, Scheme for focused classification and refinement of the BC-core region (**a**) or lobe C region of the TFIID-IIA–SCP structure (**b**) (see Methods). **c**, **d**, Fourier shell correlation plots (**c**) and local resolution estimations (**d**) of the BC-core and lobe C maps, corresponding to the final structures shown in **a** and **b**, respectively. **e**, Two-dimensional projections of the refined maps for the full TFIID-IIA–SCP structure (left), locally refined BC-core map (middle), and locally refined lobe C map (right). The maps used to calculate the projections are the same as the final structures in **a**, **b**, and Extended Data Fig. 1b, except that all have been low-pass filtered to 10 Å before calculating projections. **f**, Three-dimensional classification of 56,457 particles into two classes (solid blue and transparent green), following focused alignment to the lobe C region of the structure. The resulting classes have been superposed through their lobe C densities to illustrate the flexibility of lobe B and the upstream region of promoter DNA relative to lobe C and the downstream promoter region. The magnitude of motion within lobe A1 (20 Å) is indicated.



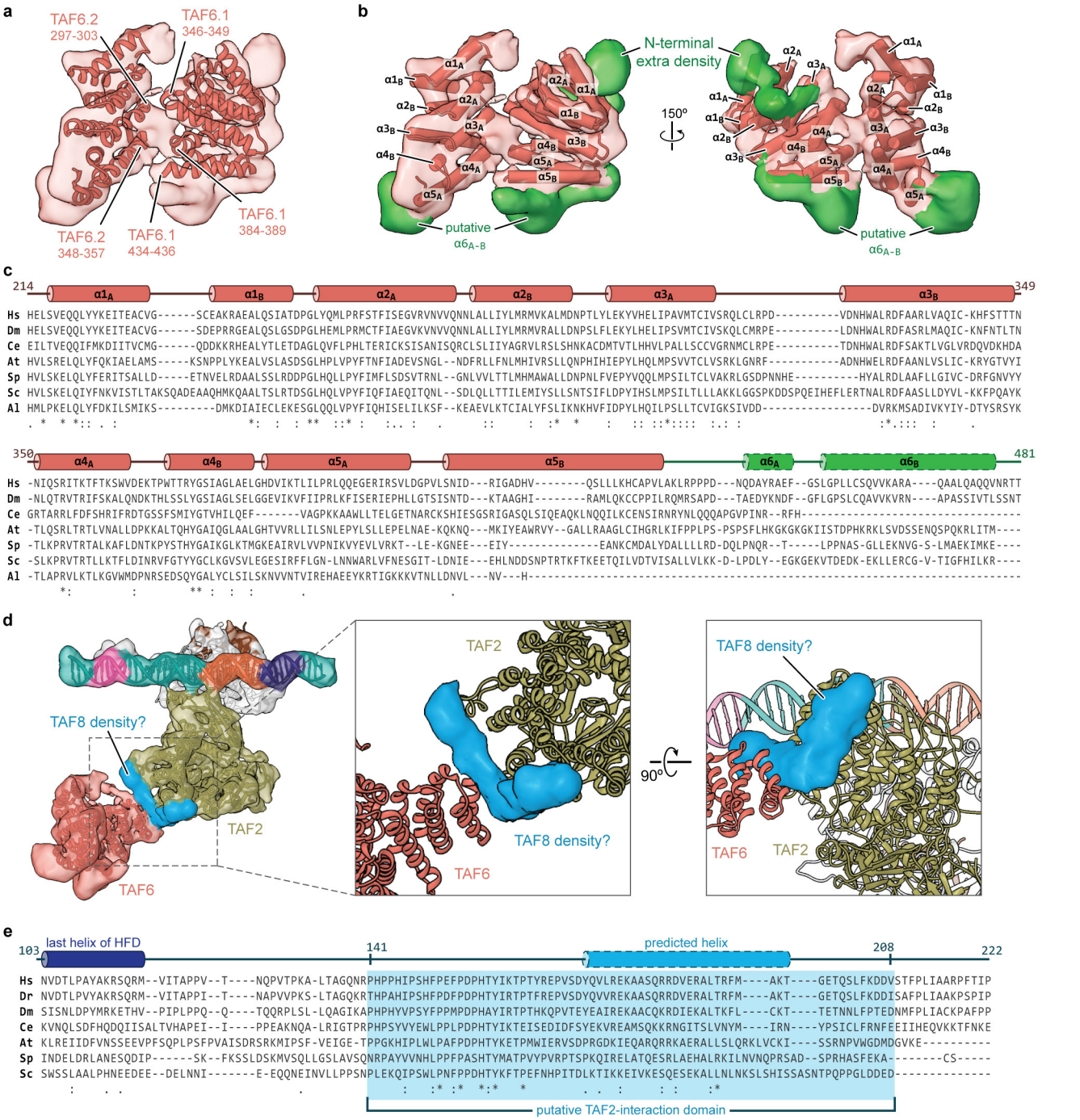
Extended Data Figure 3 | Modelling of TBP, TFIIA, and promoter DNA into the cryo-EM density. a, Previously published reconstructions of TFIID-IIA–SCP in the rearranged state (left; EMDB code 2282) and of free TFIID in the canonical state (right; EMDB code 2287)[33]. For the former, the densities for TFIIA (orange) and TBP (red) are assigned on the basis of the superposition with the TFIID-IIA–SCP structure from our present study. **b**, Close-up view of the TBP–TFIIA–TATA module density and fitted structures. The termini of the TBP structure and the three subunits (α, β, and γ) within the TFIIA structure are indicated with circles. In the cell, the α- and β-subunits of TFIIA are translated as a single polypeptide and then are post-translationally cleaved. The location of the long stretch of residues spanning the region between the structured parts of TFIIAα and TFIIAβ (TFIIAαβ 52–329) is indicated as a dashed line. Note that only 34 of the residues within this flexible loop (52–58 and 303–329) are included in the TFIIA construct used for this study. Mutational analysis in yeast has shown that mutation of an isoleucine residue (I23 in humans, I27 in yeast; represented in green spheres) to lysine at the tip of the TFIIA four-helix bundle disrupts the interaction between TFIID and TFIIA[35]. **c**, Mapping of the MPE.Fe(II) cleavage pattern for SCP DNA bound to TFIID-IIA, on the basis of data published in ref. 33. **d**, Mapping of the downstream core element (DCE)[9] sequence onto the SCP DNA within the TFIID-IIA–SCP structure from our present study.



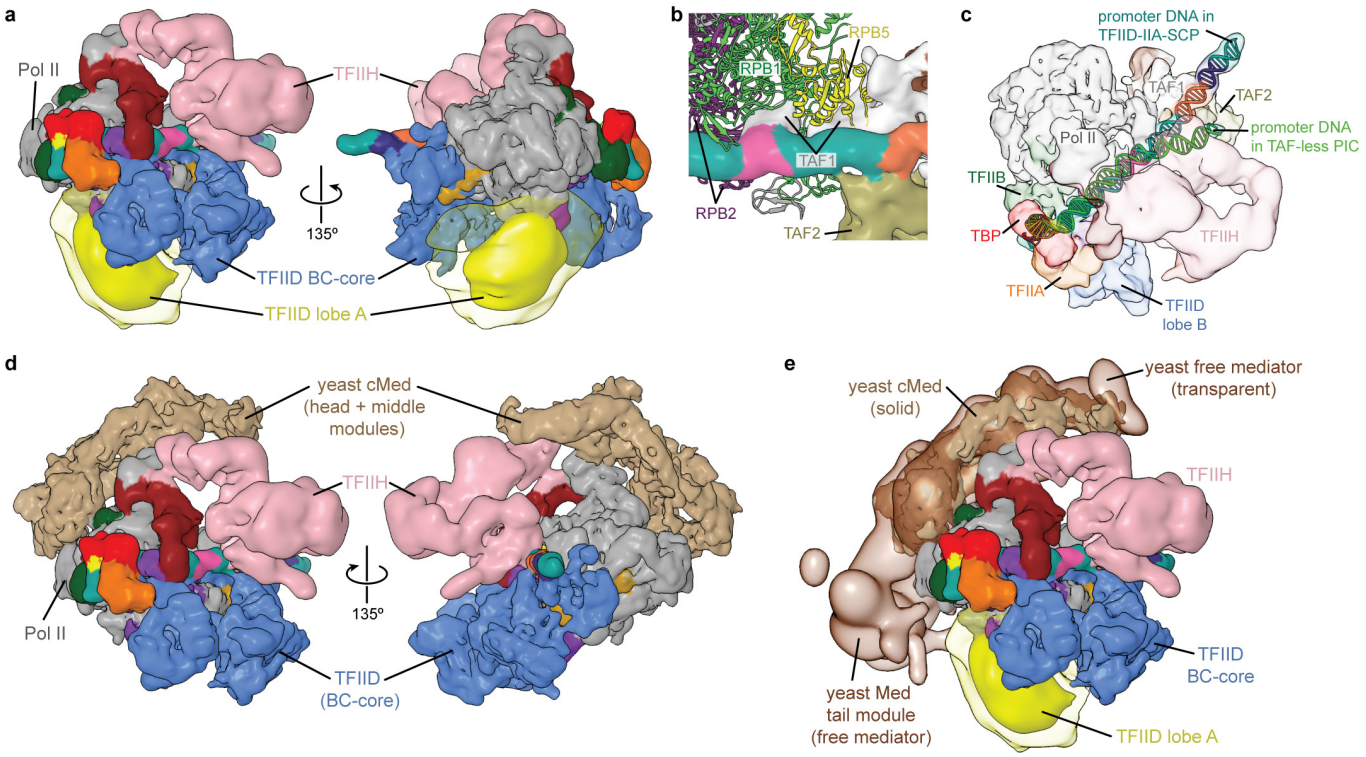
Extended Data Figure 4 | Structural modelling and conservation of the TAF1 promoter-binding domains. a, TAF1 WH domain (grey) in complex with promoter DNA (cyan) superposed with the DNA-binding WH domain of the transcription factor E2F4 (PDB accession number 1CF7, magenta) in complex with its cognate DNA, with the alignment based on the protein (left) or DNA (right) components. **b**, Sequence alignment and secondary structure map of the TAF1 WH domain, used to calculate the conservation scores depicted in Fig. 2c (Hs, *Homo sapiens*; Dr, *Danio rerio*; Dm, *Drosophila melanogaster*; Ce, *Caenorhabditis elegans*; At, *Arabidopsis thaliana*; Sp, *Schizosaccharomyces pombe*; Sc, *Saccharomyces cerevisiae*). The conserved positively charged residues that are in close proximity to the promoter DNA within the docked structure (K818, R864, K865, K868, and R875) are highlighted in pink. Numbering is based on the human sequence. **c**, Sequence alignment of a region of the TAF1 DUF3591 corresponding to the internal segment that is missing from the crystal structure and neighbouring residues. The putative Inr-binding domain (1009–1061) within this segment is highlighted in blue. Numbering is based on the human sequence, and abbreviations are the same as in **a**. **d**, Three-dimensional structure prediction for the putative TAF1 Inr-binding domain output by the I-TASSER server[67]. On the left, the residues are coloured in rainbow from N to C termini, with the terminal residues indicated. On the right, the modelling confidence is depicted in terms of the ResQ score (ribbon colour) and *B*-factor estimation (ribbon thickness) output by I-TASSER[67], with high confidence regions represented by thinner blue ribbon and low-confidence regions represented with thicker red ribbon. **e**, Secondary structure prediction for the sequence modelled in **d** (H, helix; C, coil).



Extended Data Figure 5 | Structural modelling and conservation of TAF2 APD. a, Structural arrangement of domains (D1–D4) within the TAF2 APD (bottom) compared with that of human ERAP1 (top, PDB accession number 2YD0)[39], a member of the M1 family of aminopeptidases to which TAF2 shares homology. **b**, Domain arrangement of TAF2, including the four subdomains of the APD (D1–D4), and the C-terminal intrinsically disordered region (IDR). **c**, Rigid-body docking of the best-conserved domains (D1 and D2) of the homologous human ERAP1 confirms the identity of this density. **d**, Segmented densities and fitted structures for the four subdomains (D1–D4) of the TAF2 APD. **e**, Sequence alignment and secondary structure map for the putative DNA-binding regions within domain 3 of the TAF2 APD (species abbreviations are the same as in Extended Data Fig. 4b). Conserved residues that are in close proximity to the DNA within the docked structure are highlighted in pink. The stretch that is depicted as a dashed line shares low sequence similarity with known M1 aminopeptidases. Numbering is based on the human sequence.


a
TAF6.2
297-303
TAF6.1
346-349
TAF6.1
384-389
TAF6.2
348-357
TAF6.1
434-436
b
N-terminal
extra density
150°
putative
α6A-B
putative
α6A-B
c
214
349
Hs HELSVEQQLYYKEITEACVG------SCEAKRAEALQSIATDPGLYQMLPRFSTFISEGVRVNVVQNNLALLIYLMRMVKALMDNPTLYLEKYVHELIPAVMTCIVSRQLCLRPD----------VDNHWALRDFAARLVAQIC-KHFSTTTN
Dm HELSVEQQLYYKEITEACVG------SDEPRRGEALQSLGSDPGLHEMLPRMCTFIAEGVKVNVVQNNLALLIYLMRMVRALLDNPSLFLEKYLHELIPSVMTCIVSKQLCMRPE----------LDNHWALRDFASRLMAQIC-KNFNTLTN
Ce EILTVEQQIFMKDIITVCMG------QDDKKRHEALYTLETDAGLQVFLPHLTERICKSISANISQRCLSLIIYAGRVLRSLSHNKACDMTVTLHHVLPALLSCCVGRNMCLRPE----------TDNHWALRDFSAKTLVGLVRDQVDKHDA
At HVLSRELQLYFQKIAELAMS-----KSNPPLYKEALVSLASDSGLHPLVPYFTNFIADEVSNGL--NDFRLLFNLMHIVRSLLQNPHIHIEPYLHQLMPSVVTCLVSRKLGNRF----------ADNHWELRDFAANLVSLIC-KRYGTVYI
Sp HVLSKELQLYFERITSALLD-----ETNVELRDAALSSLRDDPGLHQLLPYFIMFLSDSVTRNL--GNLVVLTTLMHMAWALLDNPNLFVEPYVQQLMPSILTCLVAKRLGSDPNNHE----------HYALRDLAAFLLGIVC-DRFGNVYY
Sc HVLSKELQIYFNKVISTLTAKSQADEAAQHMKQAALTSLRTDSGLHQLVPYFIQFIAEQITQNL--SDLQLLTTILEMIYSLLSNTSIFLDPYIHSLMPSILTLLLAKKLGGSPKDDSPQEIHEFLERTNALRDFAASLLDYVL-KKFPQAYK
Al HMLPKELQLYFDKILSMIKS---------DMKDIAIECLEKESGLQQLVPYFIQHISELILKSF--KEAEVLKTCIALYFSLIKNKHVFIDPYLHQILPSLLTCVIGKSIVDD----------DVRKMSADIVKYIY-DTYSRSYK
350
481
Hs -NIQSRITKTFTKSWVDEKTPWTTRYGSIAGLAELGHDVIKTLILPRLQQEGERIRSVLDGPVLSNID---RIGADHV--------QSLLLKHCAPVLAKLRPPPD---NQDAYRAEF--GSLGPLLCSQVVKARA----QAALQAQQVNRTT
Dm -NLQTRVTRIFSKALQNDKTHLSSLYGSIAGLSELGGEVIKVFIIPRLKFISERIEPHLLGTSISNTD---KTAAGHI--------RAMLQKCCPPILRQMRSAPD---TAEDYKNDF--GFLGPSLCQAVVKVRN----APASSIVTLSSNT
Ce GRTARRLFDFSHRIFRDTGSSFSMIYGTVHILQEF-------VAGPKKAAWLLTELGETNARCKSHIESGSRIGASQLSIQEAQKLNQQILKCENSIRNRYNLQQQAPGVPINR--RFH----------
At -TLQSRLTRTLVNALLDPKKALTQHYGAIQGLAALGHTVVRLLILSNLEPYLSLLEPELNAE-KQKNQ---MKIYEAWRVY--GALLRAAGLCIHGRLKIFPPLPS-PSPSFLHKGKGKGKIISTDPHKRKLSVDSSENQSPQKRLITM----
Sp -TLKPRVTRTALKAFLDNTKPYSTHYGAIKGLKTMGKEAIRVLVVPNIKVYEVLVRKT--LE-KGNEE---EIY----------EANKCMDALYDALLLLRD-DQLPNQR--T-----LPPNAS-GLLEKNVG-S-LMAEKIMKE----
Sc -SLKPRVTRTLLKTFLDINRVFGTYYGCLKGVSVLEGESIRFFLGN-LNNWARLVFNESGIT-LDNIE---EHLNDDSNPTRTKFTKEETQILVDTVISALLVLKK-D-LPDLY--EGKGEKVTDEDK-EKLLERCG-V-TIGFHILKR----
Al -TLAPRVLKTLKGVWMDPNRSEDSQYGALYCLSILSKNVVNTVIREHAEEYKRTIGKKKVTNLLDNVL---NV---H----------

Extended Data Figure 6 | Structural modelling and conservation of TAF6 and putative TAF8 density. a, Cryo-EM density of the TAF6 dimer with fitted homology models. Putative regions involved in the homodimer interface are labelled. **b**, Organization of α-helices within the human TAF6 HEAT-like repeat and unaccounted density (green) around the TAF6 homodimer. **c**, Sequence alignment and secondary structure map of the TAF6 HEAT repeat domain (species abbreviations are the same as in Extended Data Fig. 4b, except that Al is *A. locustae*). The green region indicates the region that is unmodelled in our structure, with the two predicted C-terminal helices outlined with dashes. Numbering is based on the human sequence. **d**, Unaccounted density indicative of two α-helices, located between domain 4 of the TAF2 APD and one copy of the TAF6 HEAT domain, which we attribute to TAF8. **e**, Sequence alignment of a putative TAF2-interaction domain within TAF8 (species abbreviations are the same as in Extended Data Fig. 4b). The last helix of the structurally determined histone fold domain of TAF8 is depicted in dark blue, while the 26 residue stretch that is predicted to be α-helical is shown in light blue with dashed outline. Secondary structure prediction was performed with PSI-PRED[71].



Extended Data Figure 7 | Modelling of the TFIID-based PIC. a, TFIID-based PIC model from Fig. 4, with the density for lobe A2 density (yellow) low-pass filtered to 16 Å and displayed at two different intensity thresholds (lower threshold in transparency). Both thresholds are lower than that used to display the density for the promoter-bound BC-core of TFIID. **b**, Close-up view of putative interactions between RPB1, -2, and -5 of Pol II and TAF1 of TFIID. **c**, Comparison of the paths of the promoter DNA within the TFIID-IIA–SCP and TAF-less PIC structures. The promoter DNA from the TFIID-IIA–SCP structure is coloured as in Fig. 1, and the promoter DNA from the TAF-less PIC is coloured in green. View is from the top of the model, relative to **a**. **d**, Docking of the core mediator coactivator complex (cMed, EMDB code 2786)[46], including the mediator head and middle modules, onto the TFIID-based PIC, on the basis of the structure of a cMED-bound initial transcribing complex. **e**, Docking of the free yeast mediator complex (brown transparency, EMDB code 2634)[47] on the basis of alignment with the core mediator shown in **c**. Lobe A2 of TFIID (yellow) is depicted similarly as in **a**.

Extended Data Table 1 | Summary of TFIID subunits

TFIID subunit	Length (residues)	M.M. (kDa)	Expected # of copies	Structured domains*	Span (residues)	M.M. (kDa)
TBP	339	38	1	DNA binding[a]	159-338	20
TAF1	1872	213	1	TAND[b]	26-87	6
				DUF3591[a]	600-1109	59
				Double bromo[b]	1359-1625	31
TAF2	1119	137	1	Aminopeptidase[a]	18-975	110
TAF3	929	104	1	Histone fold[b]	5-89	10
				Plant homeo[b]	847-921	9
TAF4	1085	110	2	TAFH[b,c]	582-678	11
				Histone fold[b,c]	870-943	9
TAF5	800	87	2	N-terminal 1[b,c]	91-124	4
				N-terminal 2[b,c]	194-340	18
				WD40 repeat[b,c]	460-739	31
TAF6	677	73	2	Histone fold[b,c]	8-77	8
				HEAT-like repeat[a,c]	212-436	25
TAF7	349	40	1	TAF1-interacting[a]	11-154	17
TAF8	310	34	1	Histone fold[b]	28-120	11
TAF9	264	29	2	Histone fold[b,c]	13-80	8
TAF10	218	22	1	Histone fold[b]	113-212	11
TAF11	211	23	1	Histone fold[b]	113-201	10
TAF12	161	18	2	Histone fold[b,c]	57-128	9
TAF13	124	14	1	Histone fold[b]	31-75	5

Subset of structured domains†	Total M.M. ‡ (kDa)
(a) Fitted domains	260
(b) Unassigned domains	290
(c) Present in the recombinant 5TAF	250

MM, molecular mass calculated from amino-acid sequence.
*'Structured domains' indicates domains that have a known structure or are predicted to be structured by sequence homology.
†The structured domains constituting each subset are indicated by superscripted letters in the larger table above, corresponding to the letter label of that subset ('a', 'b', or 'c'). Note that each domain is included in either the fitted domains subset or unassigned domains subset, on the basis of whether or not they have been modelled in the present study, respectively. Additionally, domains present in the recombinant 5TAF complex[32] constitute subset 'c'.
‡Total molecular mass for domain subsets corresponds to the total mass of structured domains and takes into account the expected number of copies for each corresponding subunit.