Near-atomic resolution visualization of human transcription promoter opening

Yuan He[1,2], **Chunli Yan**[3], **Jie Fang**[4], **Carla Inouye**[5], **Robert Tjian**[4,5,6], **Ivaylo Ivanov**[3], and **Eva Nogales**[1,4,6]

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

[2]Department of Molecular Biosciences, Northwestern University, Evanston, Illinois 60208, USA.

[3]Department of Chemistry, Center for Diagnostics and Therapeutics, Georgia State University, Atlanta, Georgia 30302, USA.

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

[5]Li Ka Shing Center for Biomedical and Health Sciences, University of California, Berkeley, California 94720, USA.

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

Abstract

In eukaryotic transcription initiation, a large multi-subunit pre-initiation complex (PIC) that assembles at the core promoter is required for the opening of the duplex DNA and identification of the start site for transcription by RNA polymerase II. Here we use cryo-electron microscropy (cryo-EM) to determine near-atomic resolution structures of the human PIC in a closed state (engaged with duplex DNA), an open state (engaged with a transcription bubble), and an initially transcribing complex (containing six base pairs of DNA–RNA hybrid). Our studies provide structures for previously uncharacterized components of the PIC, such as TFIIE and TFIIH, and segments of TFIIA, TFIIB and TFIIF. Comparison of the different structures reveals the sequential conformational changes that accompany the transition from each state to the next throughout the

transcription initiation process. This analysis illustrates the key role of TFIIB in transcription bubble stabilization and provides strong structural support for a translocase activity of XPB.

Eukaryotic gene transcription is tightly controlled during the initiation stage, when RNA polymerase II (Pol II) and the general transcription factors (GTFs) (TFIID, TFIIA, TFIIB, TFIIE, TFIIF and TFIIH) assemble at the promoter into a PIC[1–3]. Recruitment of TFIID, TFIIA and TFIIB to the DNA facilitates correct loading of the Pol II–TFIIF complex, followed by engagement of TFIIE and TFIIH[4].

The initial closed promoter complex (CC) must transition into an open complex (OC), where the melted single-stranded template DNA is inserted into the active site and Pol II locates the transcription start site (TSS). This transient OC is then converted into an initial transcribing complex (ITC), where the first phosphodiester bond forms and messenger RNA starts to be synthesized[5,6]. Following abortive cycles of synthesis of short RNA products[7], Pol II eventually clears the promoter and a stable elongation complex (EC) forms.

Crystallographic structures and crosslinking data have led to models of partial PIC assemblies for *Saccharomyces cerevisiae*[8–10]. Cryo-EM studies of human[11] and yeast[12,13] PIC complexes have described larger assemblies. Recently, we combined the cryo-EM structures of human TFIID–TFIIA–DNA and a TBP-based PIC to generate a model of the full TFIID-based structure that describes how TFIID recognizes the core promoter and contributes to PIC recruitment[14]. However, the atomic resolution description of this complex machinery and its transitions between the CC, OC, ITC and EC states are yet to be determined.

Here we describe cryo-EM structures at 3.9–8.6 Å resolution leading to near-atomic models of the human TBP-based PIC throughout transcription initiation. Comparisons between states provide new mechanistic insights into the processes of DNA engagement, promoter melting and transcription bubble stabilization.

Visualization of human initiation complexes

We used our previously developed system for assembly and purification of DNA-bound human pre-initiation complexes, in which we used TBP in place of TFIID[11]. We included TFIIS because of its potential role in PIC formation, in addition to its role in elongation[15]. Given that TFIIS can stimulate the weak RNA nuclease activity of Pol II[16,17], we also analysed the ITC structure without TFIIS (denoted hereafter as ITC(–IIS)). In order to mimic the CC, OC and ITC states, we used three different nucleic acid scaffolds (Fig. 1a). This strategy allowed us to visualize the changes in the PIC, from its assembly on duplex DNA, to one preceding clearance of the promoter. Our purified PICs were functionally active (Extended Data Fig. 1). The overall architecture of the four states analysed (Fig. 1b) resembles our reconstructions of CC published previously[11,14].

Local resolution estimation and sorting procedures (Extended data Figs 2, 3) showed that the TFIIH attachment to the rest of the PIC has some flexibility, thus limiting the resolution of the full PIC owing to slight misalignment. To overcome this limitation we also obtained

reconstructions of the 'PIC core', in the absence of TFIIH (Fig. 1c). The density maps for the full PIC and the core PIC were reconstructed using the same data set for each state, generated by automatically picking particles from images of samples that included all GTFs, but for which some of the complexes were incomplete. A different strategy was used for the core CC (see Extended Data Fig. 2 versus Extended Data Fig. 4), because for the CC sample, TFIIE was often not present in complexes lacking TFIIH (in agreement with a cooperative mode of incorporation for TFIIE and TFIIH into the PIC[11]). Instead, a local focused refinement of the CC core was conducted for the particles used to reconstruct the full CC (Extended data Fig. 2).

The global resolution for the core OC, ITC and ITC(−IIS) states is ~3.9 Å, and for the CC core is 5.4 Å. Peripheral GTFs are more flexible and at a slightly lower resolution, ranging from 4.5 Å to 7.5 Å. However, Pol II and most of TFIIB were determined beyond 4.5 Å resolution, approaching 3 Å in certain regions (Fig. 2, Extended Data Fig. 4 and Supplementary Videos 1, 2). For the rest of the complex, secondary structure was clearly visible throughout, allowing homology modelling and flexible fitting to eventually produce near-atomic models for the PIC at each step (Fig. 1d and Extended Data Fig. 5). The nucleic acids are clearly visualized, and separation between bases is apparent close to the active site (Fig. 2b and Supplementary Video 2). Interestingly, there is no density for RNA in the presence of TFIIS (ITC), whereas the RNA is clearly visible in the absence of TFIIS (ITC(−IIS)) (Fig. 2b). This result indicates that TFIIS may have changed RNA affinity for the template DNA, caused backtracking of Pol II, or activated Pol II nuclease activity.

Revised structural models for core PIC components

We traced the entire length of TFIIB in the OC, ITC and ITC(−IIS) reconstructions, and most of it in the CC structure (Fig. 3a and Extended Data Fig. 5a). Instead of the TFIIB linker helix described in crystal structures of yeast Pol II–TFIIB[8,18], we observed an extended loop directly contacting the single-stranded non-template DNA. This segment is the least conserved region in TFIIB and may represent a functional departure between yeast and humans. Alternatively, the structure of this region of TFIIB within our PIC complexes could be a result of the interaction with its natural ligand, an initiation DNA bubble, which was absent from the crystallographic studies. In fact, the entire path of the B-linker is slightly different between the human and yeast structures (Fig. 3a).

Our structures revealed an extended interface between the N-terminal lobe of TBP and residues 307–332 in TFIIAβ (Fig. 3b, Extended Data Fig. 5b). This interaction is in agreement with mutagenesis studies[19,20], and was partially captured in a structure of yeast TBP–TFIIA–TATA-box[21]. This conserved surface on TBP is a site for association with transcription activators or repressors[22,23], and has been shown to be critical for releasing the inhibitory effect of TAF1 within TFIID, possibly through binding to TFIIA[24].

The RPB1 trigger loop has been observed in crystallographic structures of the yeast Pol II in either a closed 'on' state, engaged with the last-added nucleotide[25], or an open 'off' state[26]. In all of the initiation states described here, the trigger loop adopts an 'off' state (Fig. 3c, Extended Data Fig. 5c).

Consistent with previous models of TFIIF within the PIC[11,27,28], the dimerization domain of its RAP30 and RAP74 subunits is located near the RPB2 lobe and RPB9 jaw of Pol II, whereas the C-terminal winged-helix (WH) domain in RAP30 directly contacts the BREd (downstream TFIIB recognition element) DNA, directly above the RPB2 protrusion (Fig. 3d, and Extended Data Fig. 5e). We modelled the RAP74 α1 helix, an additional α2 helix, and the linker between α1 and the dimerization domain, following a continuous region of density in our maps. Residues 119–175 in RAP30, which constitute the linker between the dimerization and WH domains, were modelled *de novo* for the four states. This linker makes extensive contacts with the RPB2 external 2 and protrusion of Pol II, as well as with TFIIB, TBP, and the BREd, implying that it plays a critical role in the correct positioning of the RAP30 WH domain for stabilization of the promoter DNA. This proposal is consistent with studies in yeast showing that any truncation from the WH end of the linker is lethal for growth[28]. We could not localize the RAP74 arm domain with certainty in any of the states, indicating that this part of TFIIF is probably flexible.

Our density maps allowed us to model the major domains in TFIIE: the WH, zinc ribbon (ZR) and helix-turn-helix (HTH) domains of TFIIEα, and the WH1 and WH2 domains of TFIIEβ (Fig. 3e and Extended Data Fig. 5d). The relative positioning of the three WH domains in TFIIE follows the previous beads-on-a-string model based on XL–MS (protein crosslinking coupled with mass spectroscopy) analysis[9], but the orientation of the domains had to be revised. The first three helices in the WH domain of TFIIEα are involved in dimerization with the WH2 domain of TFIIEβ, whereas the 4th helix directly interacts with the loop at the tip of the RPB1 clamp coiled-coil in the three open promoter states. Further down, the first helix in the WH1 domain of TFIIEβ is involved in the interaction with the WH domain of RAP30. The ZR domain of TFIIEα associates with the deep groove in the RPB7 oligonucleotide/oligosaccharide-binding fold (OB fold) in the stalk. Two helices extending from both sides of the ZR were modelled following helix-like densities directly associated with the WH in TFIIEα. Additional densities near the HTH domain of TFIIEα might correspond to the C-terminal extension of TFIIEβ following the second WH, contributing to an extended dimerization interface between the two TFIIE subunits. This proposal agrees with crosslinking studies indicating that the C terminus of TFIIEβ is close to the region flanking the ZR within TFIIEα, in addition to the WH[9,29]. This C-terminal extension has been shown to be particularly important for viability[9], in agreement with a critical contribution to the dimerization between the two subunits of TFIIE.

Despite previous efforts[30–32], we still lack a detailed subunit architecture for TFIIH. By combining the three open promoter data sets (OC, ITC and ITC(–IIS)) and following a focused refinement procedure (Extended Data Fig. 6), we produced an improved cryo-EM reconstruction of DNA-bound, 'core' TFIIH (not including the mobile kinase module[11]) that allowed the clear visualization of the major and minor grooves of the TFIIH-engaged DNA, and the modelling of six out of the seven core subunits (Fig. 3f). Consistent with our previous model[11], XPD is near the Pol II stalk and TFIIE, whereas XPB is located at the other end of the horseshoe-shaped TFIIH density, engaging the downstream DNA. When comparing our XPB model and the structure of the *Sulfolobus solfataricus* SWI2/SNF2 ATPase core bound to DNA[33], the N-terminal domain of both enzymes interacts with the duplex DNA in a very similar manner (Extended Data Fig. 6b, c). However, the C-terminal

domains are rotated with respect to each other (Extended Data Fig. 6d). The rotation resembles the conformational change proposed for the SsoRad54 ATPase[34] on the basis of a FRET (fluorescence resonance energy transfer) study and the crystallographic structure of the protein in the absence of DNA[35]. In our model, the DNA-bound XPB is in a completely different conformation from that captured for SsoRad54, emphasizing the extensive conformational plasticity of these enzymes, probably related to the ATP-driven changes that result in DNA translocation.

The crystal structure of P8 and the C-terminal dimerization domain of P52 fits well into our reconstruction, forming a heterotrimer with XPB on one side of the complex (Fig. 3f). We propose that the P52 Hub-A domain (hA-D)[32] interacts with the N-terminal, TFB2C-like domain of XPB (not built into the density), whereas P8 interacts with the XPB C-terminal domain. This model is consistent with the stimulation of XPB ATPase activity by both P52 and P8 (refs 36, 37). The N-terminal, HEAT repeat domain in P52 serves as a bridge between the two ends of TFIIH. It directly interacts with P34, which in turn contacts the P44–XPD subcomplex on the other side of TFIIH. Because of the lower resolution for this region, possibly owing to flexibility, the orientations of the P34 von Willebrand A (VWA) domain and the P52 HEAT repeat are tentative. There is still unexplained density that probably corresponds to P62 and the C-terminal ring finger domains of both P34 and P44. We propose that the P62 PH domain may correspond to a flexible density observed near TFIIE for our four reconstructions of the full PIC (Extended Data Fig. 6e).

Comparison of CC structure with previous yeast models

The crystallographic structures of yeast Pol II–TFIIB[8,18], TFIIB–TBP– TATA DNA[38], and TFIIA–TBP–TATA DNA[21] can be combined to generate a model of the Pol II–TFIIB–TFIIA–TBP and DNA complex. In this 'synthetic' structure, the TBP–TFIIA–TFIIB subcomplex and Pol II clamp and stalk regions need to be rotated to fit the human CC cryo-EM map (Extended Data Fig. 7a–c). These three modules appear inter-connected: movement of TBP–TFIIA–TFIIB–DNA closer to Pol II would cause a clash of the DNA with the clamp, thereby pushing the clamp open; opening of the clamp results in tilting of the stalk and engagement of the DNA duplex. As the axis for the TBP–TFIIA– TFIIB–DNA module rotation is by the N-terminal cyclin fold of TFIIB (Extended Data Fig. 7a), we speculate that TFIIB in the crystal structure with Pol II may be in a DNA pre-engagement state. We also observe differences between our structure and a previously published yeast cryo-EM CC model[13] (Extended Data Fig. 7d), which itself differs from the synthetic model based on crystal structures. The TBP–TFIIA–TFIIB sub-complex in the yeast CC cryo-EM study is rotated away from Pol II, which does not contact the DNA. Furthermore, the clamp is closed down, as we see for our OC and ITC complexes (see below) or that observed for EC structures[26,39]. These discrepancies may reflect known differences between the yeast and human systems. Whereas the TSS is found ~30 base pairs (bp) from the TATA box in metazoans, this distance can be as long as 120 bp in yeast[40,41]. We speculate that the yeast arrangement would allow more single-stranded DNA to be accommodated by the PIC during scanning for the TSS.

Remodelling of the PIC during promoter opening

The most notable change between the core CC and OC structures, in addition to the DNA itself (see below), is the movement of the Pol II clamp and of TFIIE, which directly contacts the clamp (Fig. 4a). The TBP–TFIIA–TFIIB–TFIIF sub-complex, defining a structural unit that surrounds the upstream core promoter elements, moves slightly, maybe as a relaxation process following the release of tension after the downstream bent DNA is melted (Fig. 4a). The bridge helix does not change during bubble opening (Fig. 2b, Fig. 4b). However, the non-template DNA near the downstream bubble junction clearly tilted the Pol II fork loop 2 in the OC, which might therefore function as a sensor of the open promoter (Fig. 2b, Fig. 4b, c). This loop stays tilted in both the ITC and ITC(−IIS) states, further supporting our hypothesis (Fig. 2b). A very similar conformation has been recently reported for the bacterial ITC[42] (Extended Data Fig. 8), suggesting the functional conservation of this critical element at the initiation stage of transcription across different kingdoms.

The most significant changes during promoter opening concern the Pol II clamp and TFIIE (Fig. 4d and Extended Data Fig. 9b). In the CC, the clamp coiled-coil domain associates with the C-terminal extension of TFIIEβ. However, in the OC, as the coiled-coil moves down when the clamp closes, the connection with TFIIEβ is broken and another interface is established with the 4th helix of TFIIEα, shifting the entire WH by ~5 Å. This change probably makes the interaction with TFIIE more energetically favourable, explaining the increase in affinity of TFIIE observed in the OC and ITC compared to the CC.

In the CC, Pol II interacts directly with the promoter DNA duplex via the 3-strand β-sheet of the clamp head (Fig. 4e, Extended Data Fig. 9c) and the 2-helix bundle in RPB5, (Fig. 4f, Extended Data Fig. 9d), immediately upstream and downstream of the initiator (INR) element (and thus the TSS). In the OC, the clamp head and RPB5 still engage the DNA, but in a slightly different manner (Fig. 4e, f and Extended Data Figs 9c, d).

For the elements described, there are no significant changes between the OC and the ITC and ITC(−IIS) states (Extended Data Figs 9a–d). Notably, the CC reconstruction does not show any detectable density for the TFIIB linker, indicating that it is disordered before promoter opening (Extended Data Figs 5a, 9a). However, in the OC, ITC and ITC(−IIS) structures, the B-linker becomes stabilized by direct interaction with the clamp coiled-coil and rudder on one side, and the non-template DNA strand on the other (Fig. 4c, e and Extended Data Figs 5a, 9a). This intricate protein–DNA network likely plays a key role in stabilizing this part of the initiation bubble, which is fairly solvent accessible.

DNA transition through initiation into elongation

The full promoter DNA is clearly visible in all our reconstructions, making it possible to compare its path throughout these transcription stages (Fig. 5). The upstream DNA, which includes BREu, TATA, and BREd elements, is tightly engaged during all stages of transcription initiation by the TBP–TFIIA–TFIIB–TFIIF sub-complex (Fig. 1c, d). After the BREd, the path of the promoter in the CC deviates from B-form DNA due to contact with the clamp and RPB5 around the TSS, and further downstream by XPB within TFIIH (Fig.

5a). The effect of XPB binding can be inferred by comparing the present CC structure with our previous cryo-EM structure of core CC in the absence of TFIIH[11] (Fig. 5a). TFIIH causes both a kink at the RPB5 contact, and a slight bend of the DNA at the site of contact with XPB (Fig. 5a, b; see also Fig. 3f). The PIC conformation maintaining this TFIIH-engaged DNA state is probably under strain, primed for DNA unwinding by the ATPase activity of XPB (see below).

In the OC cryo-EM map, two continuous densities connecting the upstream and downstream duplex DNA correspond to the separated DNA strands (Fig. 5c). The template strand follows the previously characterized tunnel[18] defined by TFIIB, and the wall, rudder, and fork loops of Pol II, whereas the non-template DNA is mainly stabilized by the B-linker and the Pol II fork loops (Fig. 4c and Extended Data Fig. 9a, c). Interestingly, while we used an 11 bp mismatch DNA scaffold for the OC, from position −9 to +2 relative to the TSS (Fig. 1a), incorporation of the DNA into the complex caused a further opening of the bubble by two bps, to position −11 (Extended Data Fig. 9e).

The most notable change between the CC and OC structures concerns the DNA itself (Fig. 4a), beginning ~20 bp downstream of the TATA box, where melting is known to start[43] (Fig. 5c). The upstream DNA barely moves. The location of the 90° bend of the DNA by TBP, further stabilized by TFIIA, TFIIB and TFIIF, likely reinforces and maintains the relative position of the immobilized upstream DNA with respect to Pol II, while a force is applied by TFIIH at the downstream end of the promoter. The minor groove of the DNA contacted by RPB5 superimposes nearly perfectly for the CC and OC reconstructions, whereas the major grooves flanking this point change to a different extent (Figs 4f, 5c). In the CC, there is a bend of the DNA at the RPB5 contact point, whereas the DNA upstream from that point of contact more closely resembles a B-form configuration than previously captured in the absence of TFIIH[11] (Fig. 5a). In the OC, the double-stranded DNA between the active site and the contact with XPB has turned as it slides past the RPB5 contact (Fig. 5c, d). In the process, the kink of the DNA by RPB5 disappears and with it probably also the tension that such a state may impose on the PIC. In contrast, the DNA downstream of the RPB5 contact does not change greatly between the CC and OC states, and XPB maintains an unchanged engagement with DNA.

There are no detectable changes in the double-stranded DNA path between the OC and the ITC, either upstream or downstream of the transcription bubble. The paths of the single-stranded template DNA are more similar than those of the non-template DNA (Fig. 6). Furthermore, whereas the template DNA in both states is visible as a continuous density, there are obvious breaks in the density corresponding to the non-template strand for the ITC. We propose that such breaks in density might be caused by disorder where scrunching is occurring.

Through all the states analysed, the downstream DNA engaged by XPB is in a duplex conformation (Fig. 3f), in agreement with previous translocase models for XPB[9,11,44]. The superimposition of the CC and OC states shows that XPB must translocate by 12 bp along the DNA to position the TSS into the active site of Pol II (Fig. 5c). The same comparison reveals a net rotation of TFIIH by ~10° with respect to the core PIC between the two states

(Extended Data Fig. 3j and Supplementary Video 3). The rotational axis passes through two critical contact points: one near the TFIIE–TFIIH interface and the other by the DNA–RPB5 interaction. TFIIH is flexibly associated to the core PIC in each of the states described (Extended Data Figs 2, 3), enabling XPB to reach the downstream DNA while TFIIH maintains contact with TFIIE. As mentioned above, initial binding of TFIIH already causes a small rearrangement of the DNA (Fig. 5a) that may prepare the promoter for melting at the correct position. Pulling of TFIIH by this strained DNA results in a slightly different trajectory of motion for the CC compared to the other states (Extended Data Figs 2, 3). Once the OC state is reached, the range of motion of TFIIH changes, and its average position rotates as indicated in Extended Data Fig. 3j. Through the transition, TFIIH would push the DNA past the RPB5 contact, which thus serves as a pivot point for both DNA and TFIIH (Supplementary Video 3). Within the context of this flexible attachment of TFIIH, the preservation of the XPB–DNA contact, at a similar position with respect to the rest of the PIC for all the states analysed, is in good agreement with a translocation model in which the opening of the transcription bubble occurs by XPB threading the DNA into the Pol II cleft, past the RPB5 contact (Fig. 5b, d, Supplementary Video 3).

When we compare our ITC(–IIS) with a recent cryo-EM structure of a bovine EC[45], the RNA–DNA hybrid and the downstream DNA superimpose, but there are significant differences for the rest of the DNA (Fig. 5e). The upstream duplex DNA in the EC is collapsed by 10 bp (RNA was normalized to the same length), whereas the path is shifted by ~15 Å towards TFIIB, to the location originally occupied by the N-terminal cyclin fold. This shift would happen at the point when the growing RNA has extended to ~12–13 nucleotides[46] and starts to displace TFIIB. Breaking contact with TFIIB, the major connection of Pol II with the rest of the PIC, should thereafter allow Pol II to clear the promoter.

The single-stranded DNA follows slightly different paths for the ITC and EC, partly because of the notable difference in the upstream duplex. Before promoter clearance, extra template DNA has to be fitted within the Pol II cleft, probably resulting in scrunching in the ITC[47,48]. This scrunching could explain the differences we observe for the non-template DNA between our ITC and the EC. The conformation of fork loop 2 in the EC is identical to what we observe in the OC, ITC and ITC(–IIS) states (Figs 5f, 2b, 4b), indicating that this element, probably important for sensing the CC to OC transition by interacting with the single-stranded non-template DNA, remains tilted during the OC to ITC to EC transitions.

Our cryo-EM study, which provides nearly complete pseudo-atomic models of all the structural elements within a functional human PIC, has allowed us to define the structural transitions through the processes of DNA engagement by the CC, opening of the transcription bubble in the OC, and initiation of transcription in the ITC. The present structures constitute a comprehensive structural framework for past and future studies of the complex process of eukaryotic transcription initiation.

Methods

Data reporting

No statistical methods were used to predetermine sample size.

Transcription complex assembly and purification

TBP, TFIIA, TFIIB, TFIIE, TFIIF and TFIIS were recombinantly expressed and purified from *Escherichia coli.* Pol II and TFIIH were immunopurified from HeLa cell nuclei (that tested negative for mycoplasma) following previously established protocols[50,51]. The design of the DNA construct was based on the super core promoter[52], except that a BREu element was introduced upstream of the TATA box[38] and an EcoRI restriction enzyme site was included downstream of the INR element for purification purposes (template, 5′-ACTGGGGAATTCCATGGTCCGTAGGCACGTCTGCTCGGCT CGAGTGTTCGATCGCGACTGAGGACGAACGCGCCCCCACCCCCTTTTA TAGGCGCCCTTC-3′; non-template 1, 5′- GAAGGGCGCCTATAAAAGGGG GTGGGGGCGCGTTCGTCCTCAGTCGCGATCGAACACTCGAGCCGAG CAGACGTGCCTACGGACCATGGAATTCCCCAGT-3′). The nucleic acid scaffolds that were used to generate the PICs in the open and initial transcribing conformation were designed by modification of the promoter substrate used to form the closed PIC. 11 bp or 19 bp mismatch sequences composed of poly-T were introduced to mimic the different sizes of initiation bubbles in the open and initial transcribing conformation during promoter opening, respectively (Fig. 1, top). Though the mismatch bubble stops at −9 in the open conformation, PIC assembly further unzips two extra base pairs at the upstream end of the bubble. Mismatch bubbles that stop at −11 and −7 were tested in the ITC reconstruction. The latter scaffold did not generate a stable complex, indicating that the GTFs can open up two, but not four, base pairs in the context of the core PIC. An RNA–DNA duplex beyond 7 bp or 12 bp has been proposed to be the trigger for TFIIB release and promoter escape[8,10,18,46,53]. Thus, we introduced a 6 nucleotide (nt) RNA primer complementary to the template strand, starting at the TSS used in our study (non-template 2, 5′-GAAGGGCGCCTATAAAAGGGGGTGGGGGCGCGAAT TTTTTTTTTCGCGATCGAACACTCGAGCCGAGCAGACGTGCCTACGGA CCATGGAATTCCCCAGT-3′; non-template-3, 5′-GAAGGGCGCCTATA AAAGGGGGTGGGGGCGTTTTTTTTTTTTTTTTTTTTTCGAACACTCGA GCCGAGCAGACGTGCCTACGGACCATGGAATTCCCCAGT-3′; RNA, 5′-AGUCGC-3′). A biotin tag was engineered at the 5′ end of the template strand (Integrated DNA Technologies). The duplexed DNA was generated by annealing the template strand with equimolar amounts of single-stranded non-template DNA at a final concentration of 50 μM in water. The annealing reaction was carried out at 100 °C for 5 min and gradually cooled down to room temperature within 2 h.

PIC in the closed conformation was assembled as initially described[11], except for an additional incubation of TFIIS at a final concentration of 200 nM with the purified PIC before application to the EM grid[14]. PICs in the open and initial transcribing conformation were assembled in a similar fashion, but using the appropriate nucleic acid scaffolds instead of DNA duplex. Purified PIC complexes were crosslinked after elution by incubation with

glutaraldehyde at a final concentration of 0.05%, on ice and under very low illumination conditions, for 5 min, then immediately used for cryo-EM sample preparation.

Electron microscopy

Preparation of PIC samples for cryo-EM observation was carried out using 400 mesh C-flat grids containing 4 μm holes with 4 μm spacing (Protochips). A thin carbon film was floated onto the grid before it was plasma cleaned for 5 s using a Solarus plasma cleaner (Gatan) under a 75% argon/25% oxygen gas mix, immediately before sample deposition. An aliquot (3 μl) of the purified sample (~100 nM) was placed onto the grid and loaded into a Vitrobot (FEI) operated at 100% humidity and 4 °C. The sample was allowed to absorb for 5 min (under low illumination conditions), then was blotted for 4 s and immediate plunged into liquid ethane. The frozen grids were stored in liquid nitrogen until loaded into a Titan electron microscope (FEI) operating at 300 keV using a 626 single-tilt cryotransfer system (Gatan). Data were acquired with a K2 summit direct electron detector (Gatan) operating in counting mode at a calibrated magnification of $37,879 \times$ (1.31 Å per pixel), using a range of defocus values (from −2 to −4 μm). Between 855 and 1334 movie series for each of the cryo-EM data sets were collected using the MSI-T application of the Leginon data collection software[54]. 30-frame exposures were taken at 0.3 s per frame (9 s total exposure time), using a dose rate of 8 e– per pixel per second, corresponding to a total dose of 42 e– Å^{-2} per movie series.

Image processing

Movie frames were aligned using MotionCorr to correct for specimen motion[55]. The average of the aligned frames was used for data preprocessing within the Appion processing environment[56]. Particles were automatically selected from the aligned micrographs using a difference of Gaussians (DoG) particle picker[57]. For CC, OC, ITC and ITC(–IIS), 245,501, 328,720, 366,145, and 434,396 particles were picked from 855, 966, 1099, and 1,344 micrographs, respectively (Extended Data Figs 2–4). The contract transfer function (CTF) of each micrograph was estimated using the CTFFind3 program[58] during data collection. All three-dimensional (3D) classification and refinement steps were performed within RELION[59] (version 1.4-beta).

Three-dimensional reconstruction

For the CC, the initial set of 245,501 particles was subjected to an initial 3D classification, using the negative stain reconstruction of a similar complex[11] (in which the nucleic acid is not visible and TFIIS is absent) low-pass filtered to 60 Å as the initial reference (Extended Data Fig. 2c). Because of the use of such reference, the presence of density for the DNA and for TFIIS in the reconstruction can be used as clear signs that features are not arising owing to noise alignment to the references. One out of five classes in this classification, corresponding to 34,728 particles, was indicative of well-preserved holo-complex with sharp structural feature and was selected for further processing. This particle set was then subject to 3D auto-refinement within RELION, resulting in a reconstruction of the CC at an overall resolution of 7.2 Å (Extended Data Fig. 2b). All resolutions reported herein correspond to the gold-standard Fourier shell correlation (FSC) using the 0.143 criterion[60]. Local resolution estimation indicated that the density for TFIIH core (the kinase module is too

flexible to be visualized) was at lower resolution than the PIC core containing the rest of the proteins, probably owing to conformational heterogeneity (Extended Data Fig. 2c). Indeed, further 3D classification within RELION successfully separated the data set into two, resulting in two distinct locations of the TFIIH relative to the PIC core (Extended Data Fig. 2c). This flexibility of a major part of the complex would also likely have contributed to the limited resolution of the rest of the CC reconstruction. Therefore, a soft mask was applied around the PIC core density (that is, excluding TFIIH) during further 3D refinement within RELION, thus effectively excluding the contribution of the TFIIH signal from the final alignment rounds. This procedure resulted in an improved reconstruction of the PIC core, with an overall resolution of 5.4 Å (Extended Data Fig. 2b, c).

For the OC, the initial set of 328,720 particles was subjected to an initial 3D classification, similar to the above CC (Extended Data Fig. 3c). One out of five classes in this classification, corresponding to 59,271 particles, was indicative of a well-preserved holo-complex and was selected for further processing. This particle set was then subject to 3D auto-refinement within RELION, resulting in a reconstruction of the OC at 8.6 Å resolution (Extended Data Fig. 3b). Further 3D classification within RELION also successfully separate the data set into two, resulting in two distinct locations of TFIIH relative to the PIC core (Extended Data Fig. 3c). However, the movement trajectory of TFIIH found in the OC was different from the one found in the CC. The same initial data set was also used to derive the reconstruction of the PIC core using the negative stain reconstruction of a similar complex[11] (in which the nucleic acid is not visible and TFIIS is absent) low-pass filtered to 60 Å as the initial reference (Extended Data Fig. 4c). One out of five classes in this classification, corresponding to 79,849 particles, was indicative of well-preserved full core complex and was selected for further processing. This particle set was then subject to 3D auto-refinement followed by particle polishing procedure within RELION[61] in order to correct for individual particle motion and beam-induced radiation damage of the sample. Dose-dependent relative B-factor was estimated per frame and applied during the particle polishing step (Extended Data Fig. 4a). This resulted in a reconstruction of the OC in the absence of TFIIH at an overall resolution of 3.9 Å (Extended Data Fig. 4b). Similar procedures were taken for processing the data set for the ITC and ITC(−IIS) (Extended Data Figs 3, 4).

For the TFIIH core complex, three data sets corresponding to the OC, ITC and ITC(−IIS) were merged together following the initial 3D classification within each data set giving rise to the holo-PIC (Extended Data Fig. 6a). A similar approach to that used for deriving the PIC core within the CC was taken by applying a soft mask around the TFIIH core density during further 3D refinement within RELION, effectively excluding the contribution of core PIC signal during the alignment. This procedure resulted in an improved reconstruction of the TFIIH core, with clear observation of the major and minor grooves of the bound duplex DNA (Fig. 3f and Extended Data Fig. 6a).

Local resolution calculation was performed for all reconstructions using the 'blocres' function in the Bsoft package[62,63] (Extended Data Figs 2–4, 6a). The final density maps were automatically sharpened using the post-processing program within RELION and then filtered according to local resolution using the 'blocfilt' function within Bsoft. Volume

segmentation, automatic rigid-body docking, figure and movie generation were performed using UCSF Chimera[64].

Building initial PIC atomic models

To interpret the structural mechanism in more detail, four models of the PIC (CC, OC, ITC and ITC(–IIS)) in different functional states were built. We first generated an initial homology model of human RNA polymerase II (Pol II) in complex with TFIIB using MODELLER[65,66]. We used the sequences of the twelve Pol II subunits (RPB1 to RPB12) and TFIIB without the C-terminal cyclin fold in the crystal structure of yeast *Saccharomyces cerevisiae* Pol II–TFIIB complex (PDB: 4BBR)[10] as a template. This homology model for Pol II–TFIIB was placed as a rigid body into the density map using UCSF Chimera[64]. The RPB4 and RPB7 subunits were then replaced by the X-ray structure of human RPB4 and RPB7 heterodimer (PDB: 2C35)[67]. Analysis of the fit showed that the B-linker helix of yeast TFIIB overlapped with the DNA in the cryo-EM reconstruction and that it required to be retraced as a loop region.

The crystal structure of TBP, TFIIB and upstream core promoter DNA duplex (PDB: 1C9B)[38] was placed and rigid-body fitted into the map by aligning TFIIB to the TFIIB's N-terminal cyclin-fold domain. We then extended the model to include TFIIA, by superimposing TBP on the TBP–TFIIA–DNA crystal structure (PDB: 1NVP)[21].

The crystal structure of the human TFIIF dimerization domain (PDB: 1F3U)[68] was docked into the density ascribed to TFIIF by the lobe domain of Pol II. Our rigid body fitting of crystal structures results in a small clash between the RPB2 lobe and the RAP74 C-terminal α1 helix, a region shown to be important in both transcription initiation and elongation. RAP74 α1 helix was rigid-body adjusted to better fit the density and extra residues (residues 167–180) in the C terminus were included based on the results of secondary structure prediction[69,70]. In addition to interacting with Pol II, it is evident from the cryo-EM density that TFIIF directly contacts the BREd DNA. We previously proposed[11] that this region corresponds to the C-terminal winged-helix (WH) domain of RAP30 (PDB: 1BBY)[71], and we can now dock it with high accuracy into our improved map. TFIIF also contributes to the extra density seen linking the C-terminal WH domain and the dimerization domain near the protrusion of Pol II right below the central cleft. We built the path of RAP30, N-terminal to the WH domain on the basis of secondary structure prediction[69,70] and fitted it into the electron microscopy density. Two different protein secondary prediction servers (PSIPRED and SABLE) were used for cross-validation purposes[69,70].

To model the human TFIIE WH domains, residues 11–101, corresponding to the TFIIEα WH, were aligned with *Sulfolobus solfataricus* TFIIE WH residues 7–88 (PDB: 1Q1H)[72]. TFIIEβ WH2 residues 150–207 were aligned with *Sulfolobus tokodaii* STO12A residues 25–106 (PDB: 2D1H)[73]. The human ZR domain, residues 125–164 (PDB: 1VD4)[74], was docked near the base of the stalk domain of Pol II. The human TFIIEβ WH1 residues 75–146 (PDB: 1D8J)[75] were fitted into the region proximal to TFIIF RAP30 WH domain. The individual TFIIE WH domain structures were oriented relative to each other by rigid-body fitting into the electron microscopy density. The TFIIEα residue segments 102–124 and 165–204 and the TFIIEβ residues 205–239 were modelled on the basis of the observed

electron microscopy density and consensus secondary structure prediction[69,70]. Missing loops and the linker between the WH domains of TFIIE were built with the program Modeller[65,66].

The X-ray structure of human TFIIS domain II (PDB: 3NDQ) and NMR structure of the TFIIS zinc-finger motif (PDB: 1TFI)[76] were combined and rigid-body fit into the density for the core of Pol II. The linker between domain II and the zinc-finger motif was constructed from the yeast TFIIS structure (PDB: 1Y1V)[26] by homology modelling and fit into the density.

To model TFIIH, we used the crystal structure of the *Archaeoglobus fulgidus* XPB N-terminal domain (PDB: 2FZ4)[77] as a template to build the corresponding N-terminal portion of human XPB. The C-terminal domain of human XPB (PDB: 4ERN)[78] was then directly used to fit into our density map as a rigid body before connecting with the N-terminal part. We generated a homology model for the XPD on the basis of the *Thermoplasma acidophilum* XPD structure (PDB: 2VSF)[79]. The N-terminal helix was added by using *Sulfolobus acidocaldarius* XPD (PDB: 3CRV)[80]. The X-ray structure for Ssl1/P44 from *Saccharomyces cerevisiae* (PDB: 4WFQ)[81] was used as a template to construct the human P44 structure. We used the crystal structure of a minimal complex between Tfb5, the yeast orthologue of P8, and the C-terminal domain of Tfb2, the yeast P52 subunit of TFIIH (PDB: 3DOM)[82], as a template to build the human P52/P8 structure. The human structure of P52/P8 was then docked into a corresponding density adjacent to the C-terminal half of XPB. In order to remove the clashes between the N-terminal helix of P52 and XPB, we rigid-body adjusted the position of this helix to better fit the density. The structure of the N terminus of p52/Tfb2 is predicted to be similar to the HEAT repeats of human transportin 3. We used the human X-ray structure of transportin 4 (PDB: 4C0O)[83] as a template to model the human P52 N-terminal heat repeat. The amino acid sequence of human P52 was aligned with that of human transportin 4, initially with the PROMALS3D multiple sequences and structure alignment server[84], and then adjusted manually. The human p44 VWA domain was modelled using the X-ray structure of the p34 subunit of the TFIIH complex from the eukaryotic thermophilic fungus *Chaetomium thermophilum* (PDB: 4PN7)[85].

The upstream core promoter DNA elements within the OC, ITC and ITC-TFIIS complexes remained bound by TBP, TFIIA and TFIIB, in the same configuration as in the CC. The bent DNA model in the CC was generated using the 3D-DART online server and the 3DNA software package[86]. To model downstream DNA, the DNA–RNA hybrid structure from a previous model of the budding yeast elongation complex (PDB: 4A3F, 4A3D and 4A3I)[5] was fit into the density and adjusted. The downstream DNA was extended by adding duplex DNA in each model. Individual nucleotides in the DNA bubble were adjusted to best represent the observed electron microscopy density in this region of the cryo-EM maps. The sequences of the DNA substrates (in the electron microscopy experiments and the simulation) are shown in Fig. 1a. These initial models (for core CC, OC, ITC, ITC(−IIS), and for TFIIH) were subjected to electron microscopy flexible fitting and refinement of atomic positions as described in the next section.

Molecular dynamics flexible fitting and refinement in Phenix

To better capture the conformational changes revealed by the cryo-EM maps in different states, a molecular dynamics flexible fitting procedure (MDFF) was used to flexibly fit the models into the EM density[87]. MDFF is a method wherein external forces proportional to the electron microscopy density gradient are applied, biasing the motion of the protein atoms towards the high-density regions of the map. Additional restraining forces were also applied during fitting to prevent structural distortions and preserve the secondary structure elements. Hydrogen atoms, counterions (Na^+) and TIP3P solvent were introduced using the Tleap module in AMBER10[88]. Additionally, 100 mM NaCl concentration was introduced to mimic physiological conditions. Production runs were carried out in the NPT ensemble (1 atm and 300 K) for 10 ns for each of the four models of the PIC complex (denoted OC, ITC, ITC(−IIS) and CC). We employed MDFF in several stages with progressively higher values for the MDFF force scaling factor ξ. These were varied from 0.1 to 0.3. The XPB/DNA and XPD/P44 systems (IIH model) were simulated for 5 ns with a scaling factor of 0.1. We allowed necessary structural adjustments during MDFF fitting. All simulations were performed using the NAMD2.8 code[89] with the AMBER Parm99SB parameter set containing the basic force field for nucleic acids and proteins, as well as the refined parameters for backbone dihedrals for protein (SB) and nucleic acids dihedrals (BSC0)[90] on the Stampede supercomputer at the Texas Advanced Computing Center and the Edison supercomputer at the National Energy Research Scientific Computing Center (NERSC). MDFF flexible fitting was followed-up by refinement of atomic coordinates and atomic displacement parameters (ADP) using the Phenix package[91]. To carry out reciprocal space refinement, the electron microscopy density maps were converted to structure factors. First, we performed ten macro-cycles of real space refinement applying secondary structure restraints. Then we carried out eight macro-cycles of reciprocal space refinement in Phenix with optimization of the map/stereochemistry weight restraints and applying the MDFF-fitted structures as reference models[92]. To minimize potential overfitting, geometric statistics were monitored during the various rounds of refinements. Finally, the resulting models were validated using MOLPROBITY[93] (see data statistics in Extended Data Table 1). After refinement had converged, the models for the core CC, OC, ITC and ITC(−IIS) and the TFIIH subunits (XPB, XPD, P44, P34, and P52/P8) were separately rigid-body fitted into the respective maps of the holo-complexes (that is, with TFIIH) and the models combined to assemble the full PIC-TFIIH models.
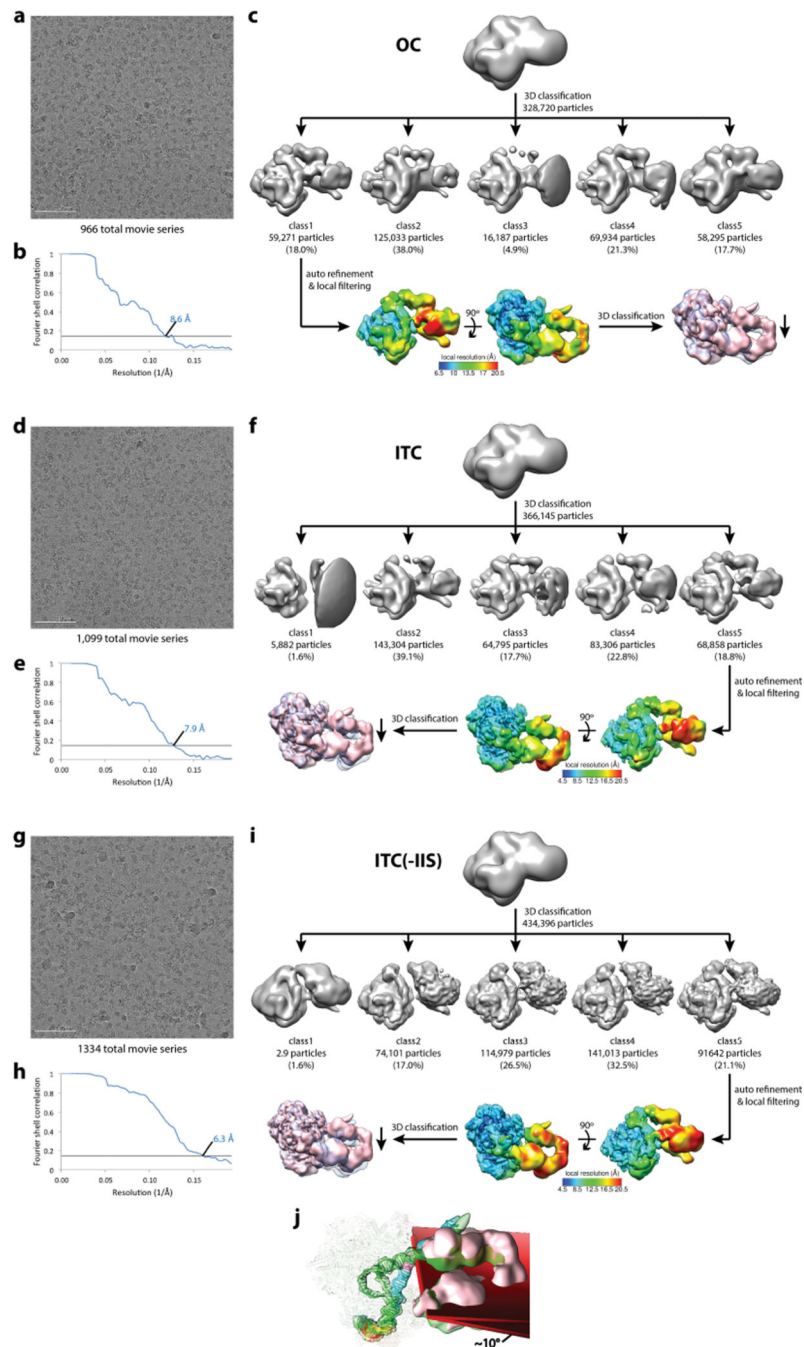
Extended Data
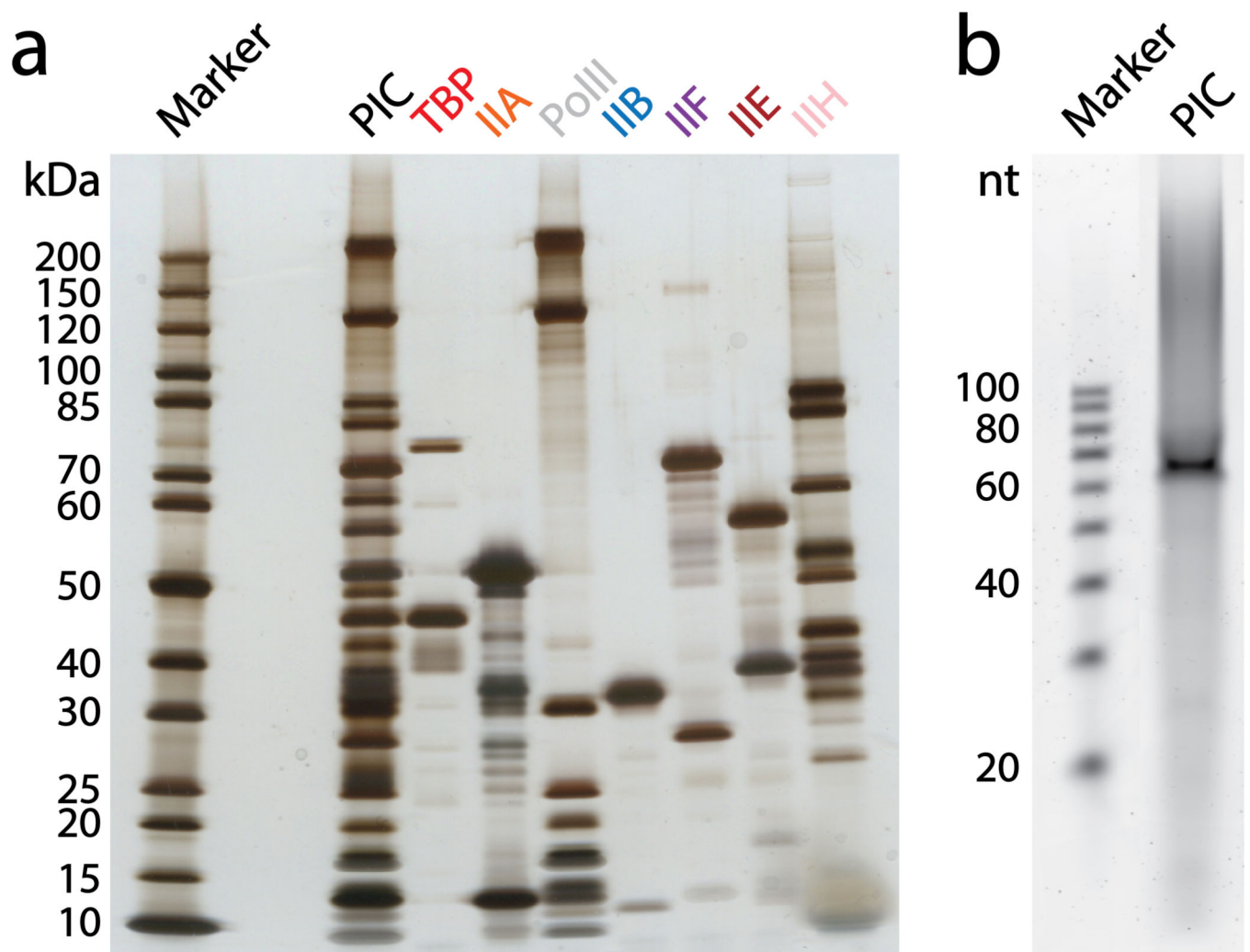
Extended Data Figure 1. Purification and activity of human PIC
a, SDS–PAGE (4–12% gradient gel followed by silver staining) of purified transcription factors and assembled PIC on a promoter DNA. **b**, The purified PIC in **a**, was ran on a 10% TBE-urea gel after supplying ribonucleoside triphosphates in a run-off reaction. kDa, kilodaltons.
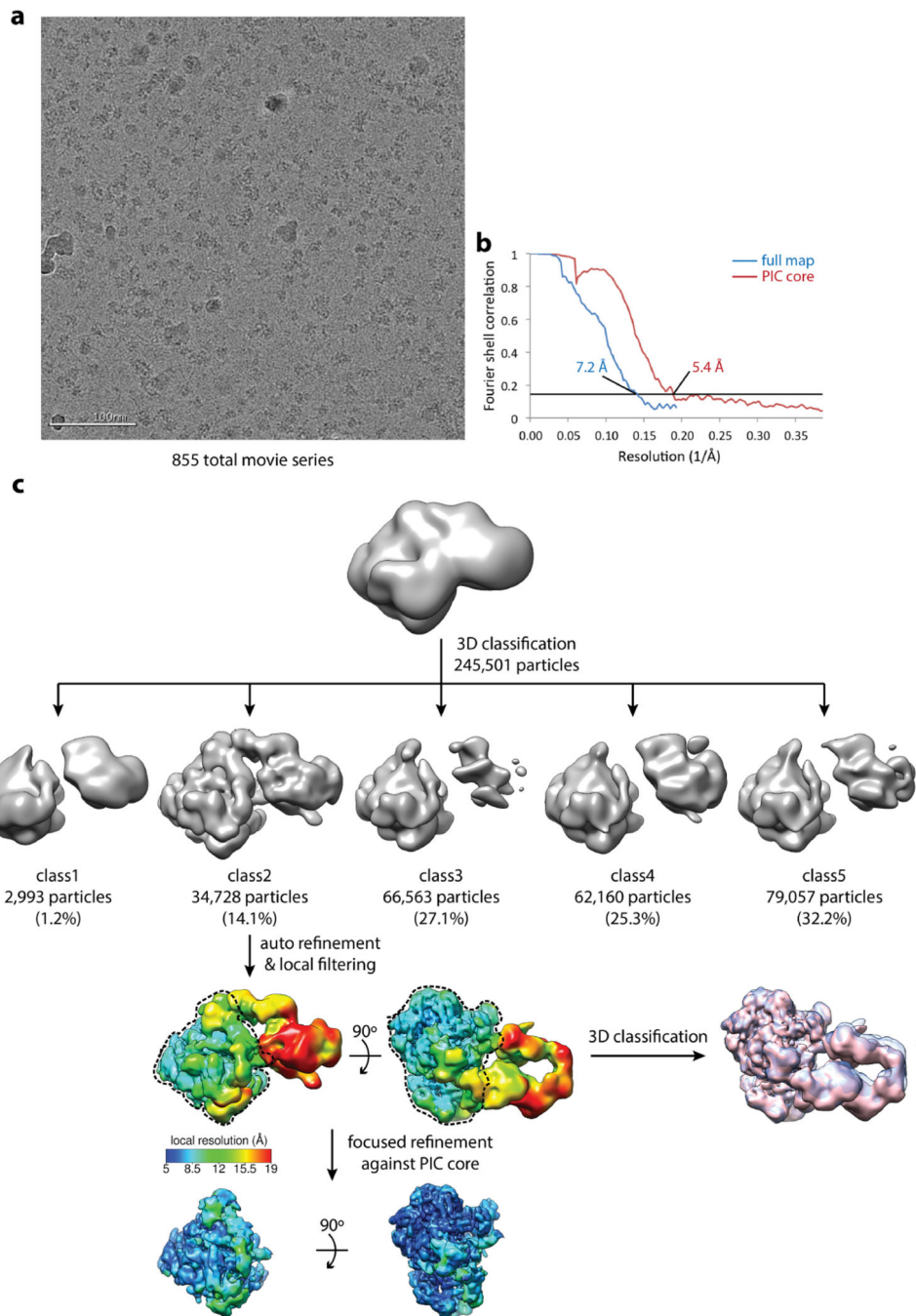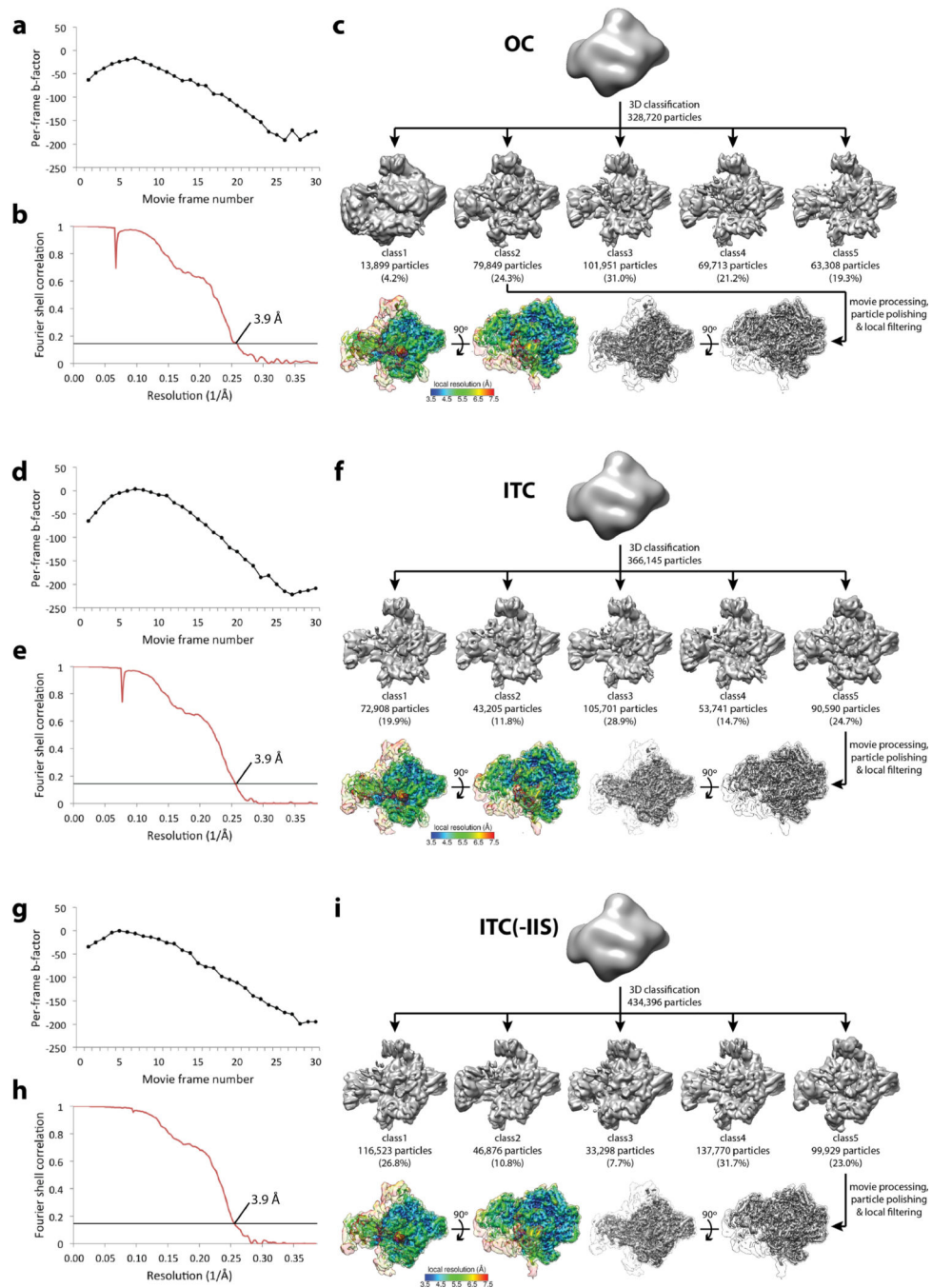
Extended Data Figure 2. Cryo-EM of the human CC
a, Representative raw micrograph. **b**, FSC curve and estimated resolution using the 0.143 criteria following the gold-standard procedure implemented in RELION for both the holo-complex and the PIC core. **c**, Refinement strategy for the holo-complex (see Methods). The local resolution estimation shows flexibility for TFIIH. Further 3D classification revealed the range of motion of TFIIH within the complex (pink and sky blue densities). Focused refinement on the PIC core (masking out TFIIH) improved alignment accuracy and improved the resolution for the core complex.

Extended Data Figure 3. Cryo-EM of the human OC/ITC/ITC(–IIS) complexes
a, d, g, Representative raw micrograph. **b, e, h**, FSC curve and estimated resolution using the 0.143 criteria following the gold-standard procedure implemented in RELION. **c, f, i**, Refinement strategy for the holo-complex (see Methods). The local resolution estimation shows flexibility for TFIIH. Further 3D classification revealed the range of motion of TFIIH within the complex (pink and blue densities). Notice that the direction and range of motion for these three states is similar. Compared to the CC, the pink densities are approximately

the same, but the direction of motion from pink to blue has changed. **j**, Movement of TFIIH with respect to the core PIC from the CC to OC state.


Extended Data Figure 4. Cryo-EM of the core human OC/ITC/ITC(–IIS) complexes
a, d, g, Dose-dependent B-factor plot from RELION. **b, e, h**, FSC curve and estimated resolution using the 0.143 criteria following the gold standard procedure implemented in RELION. **c, f, i**, Refinement strategy (see Methods). The local resolution estimation and final density map (filtered according to this local resolution) is shown for two different

views and at two different thresholds (the higher threshold allows better visualization of the highest resolution features for the more stable elements).
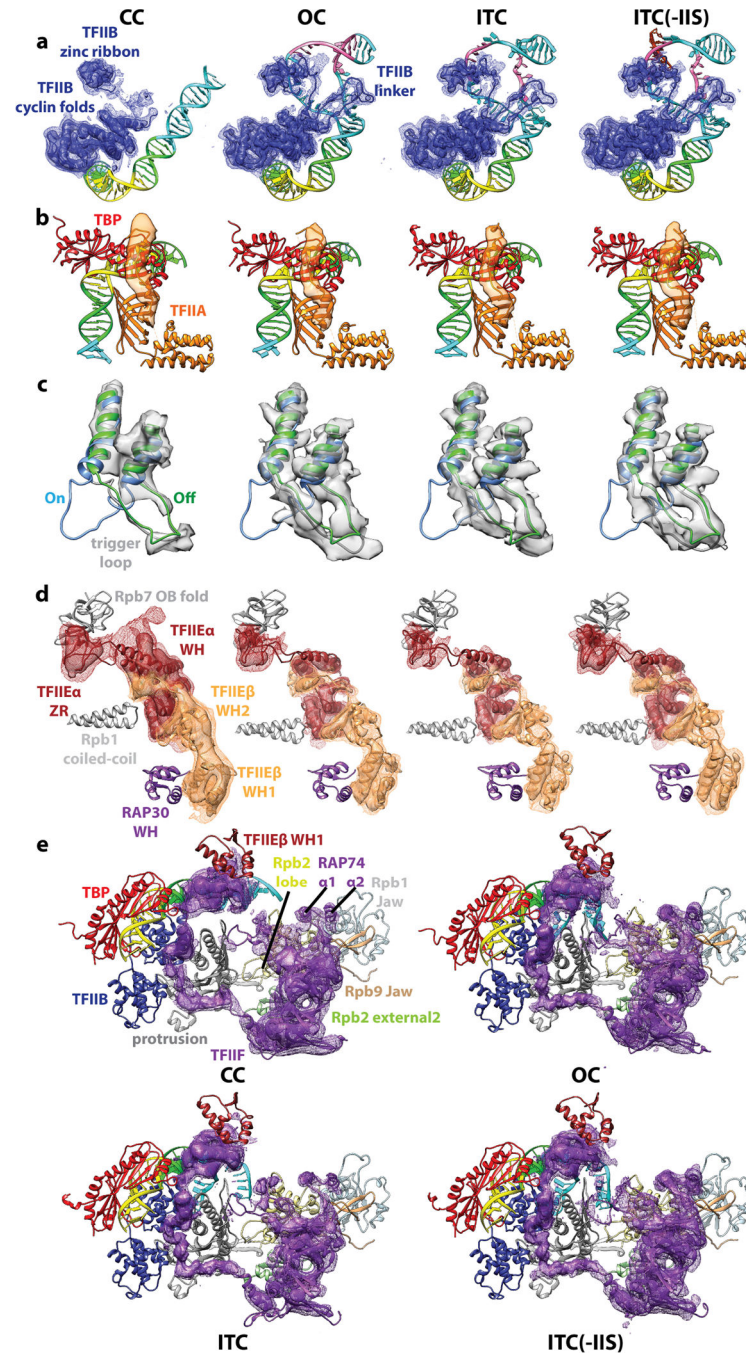

Extended Data Figure 5. Segmented cryo-EM densities and models for different regions of interest within the human core PIC in the CC, OC, ITC and ITC(–IIS) states
a, TFIIB and its interaction with DNA. **b**, Extension of TFIIA stabilized by interaction with TBP. **c**, Trigger loop in Pol II, viewed in an 'off ' state by comparison with crystallographic structures of yeast Pol II (2NVZ[25] in blue and 1Y1V[26] in green). **d**, TFIIE and its interacting partners with the PIC. **e**, TFIIF and its interacting partners with the PIC.
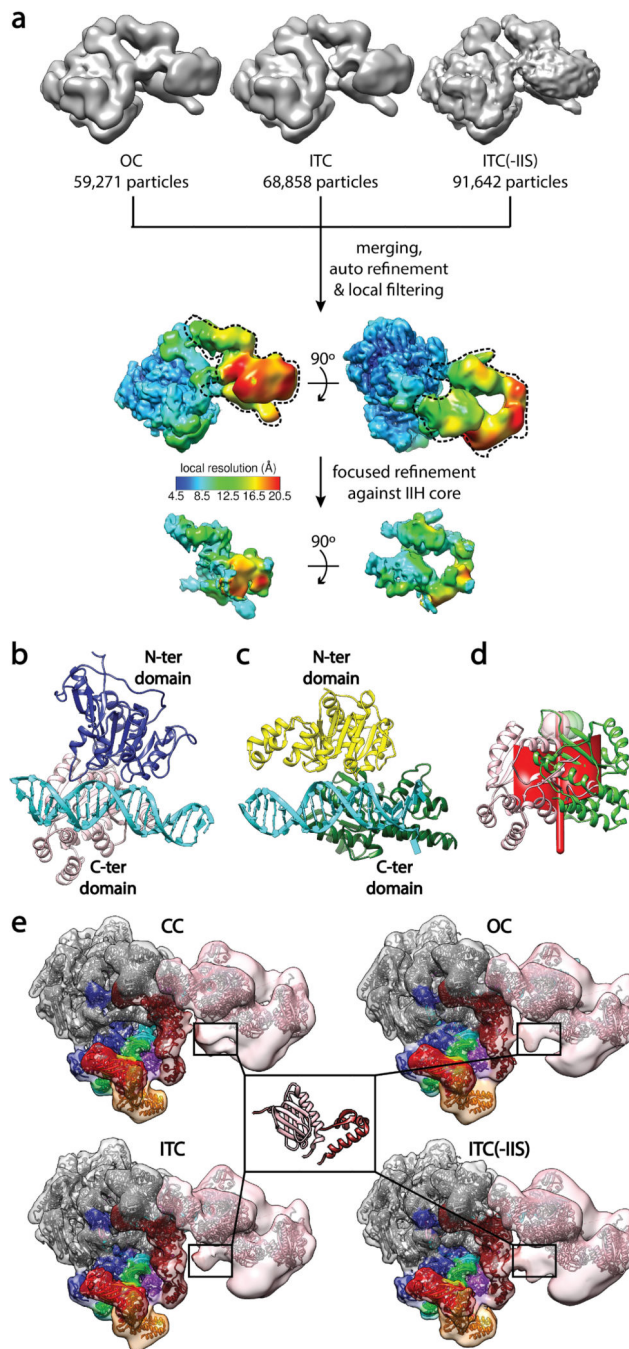
Extended Data Figure 6. Cryo-EM reconstruction and interpretation of human TFIIH
a, Refinement strategy (see Methods) for the seven subunit TFIIH core complex (the CAK subcomplex is too flexible to be visualized following averaging). The local resolution estimation for the PIC structure obtained by combining the OC, ITC and ITC(–IIS) data sets shows lower resolution for TFIIH, reflecting its flexible attachment. Focused refinement (using the mask marked by dashed lines) allowed us to improve the resolution for TFIIH. **b–d**, Comparison of human XPB and the ssoRad54 ATPase structures and their interaction with DNA. **b**, Human XPB–DNA model. N terminus is shown in navy blue, C terminus in

pink and DNA in cyan. **c**, Crystal structure of ssoRad54 ATPase (PDB: 1Z63)[33]. The N terminus is shown in yellow, the C terminus in green and the DNA in cyan. Whereas the N-terminal domains of both proteins and the DNA can be easily superimposed, the C-terminal domains are in very different position. **d**, A 123° rotation would be required to superimpose the C-terminal domains around an axis located on the first residues (top, shown in spheres) that connects to the fixed N-terminal domain. The rotation axis and planes are coloured in red. **e**, Possible location of the TFIIEα–TFIIH/p62 interface. An unassigned density in the region of proximity/contact between TFIIE and TFIIH probably corresponding to the TFIIE–TFIIH interface is marked with boxes for each reconstruction. A tentative orientation of the NMR structure for a short C-terminal segment of TFIIEα bound to the PHD domain of p62 (PDB: 2RNR)[94] is proposed in the centre for that flexible density region.
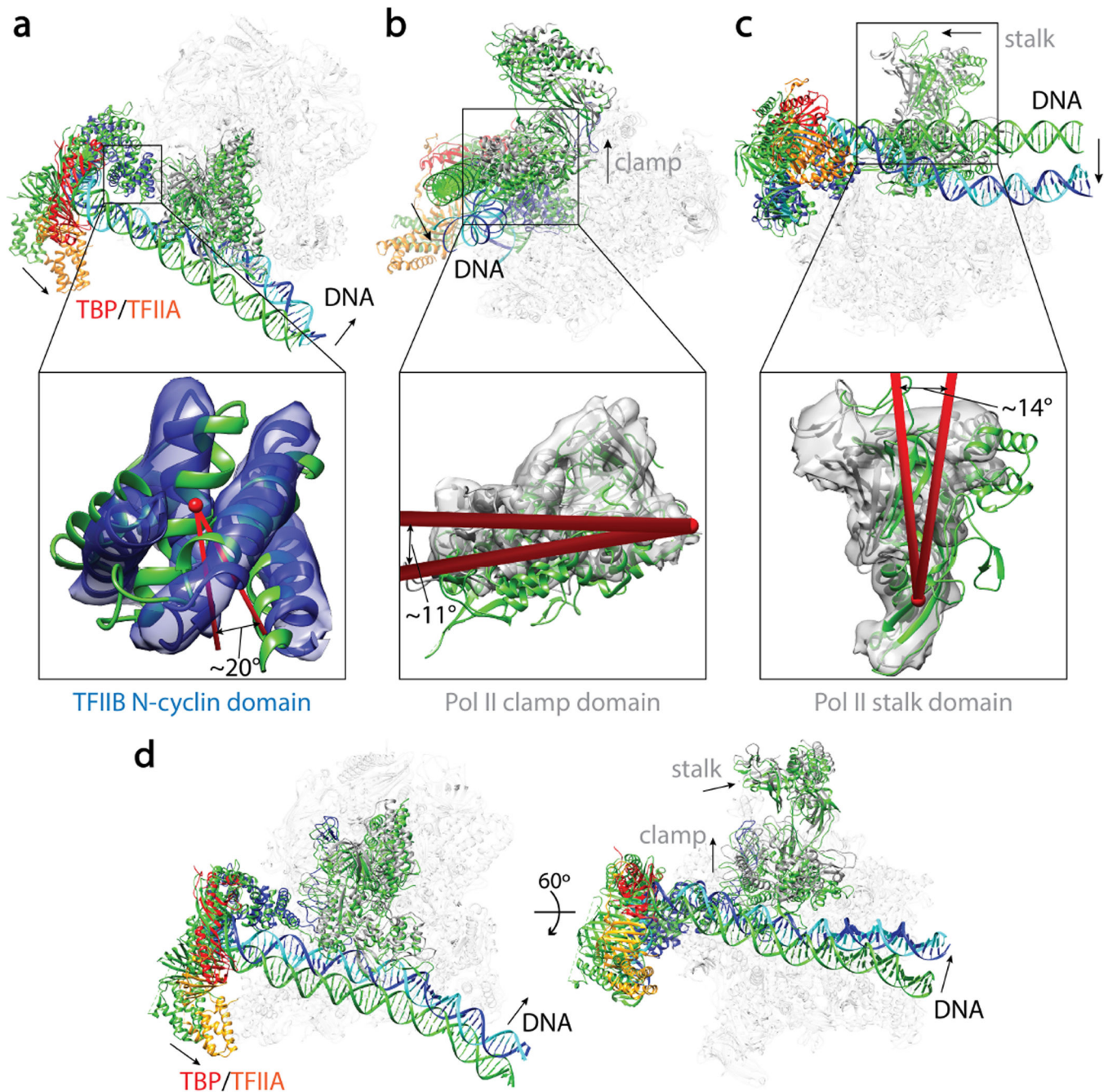
Extended Data Figure 7. Position of the TBP–IIA–IIB module and mobile elements of Pol II
a–c, Comparison between a 'synthetic' structure of the CC, generated by superimposing human TBP–TFIIA–DNA (PDB: 1NVP)[21], human TBP–TFIIB–DNA (PDB: 2C9B)[38], and, most notably, the yeast Pol II–TFIIB (PDB: 4BBR)[10] using common elements (green), and the human cryo-EM CC model (coloured). A number of elements are rotated between the two: the TBP-IIA-IIB-DNA subcomplex (**a**), the Pol II clamp (**b**), and the Pol II stalk (**c**). The rotation plane and angle are depicted in red. **d**, Comparison, shown in two different views, of a recently reported cryo-EM structure of yeast CC (green)[13] and the human CC in this study. The structures were aligned using the rigid part of Pol II (that is, excluding the

clamp and the stalk). The yeast TBP–TFIIA– TFIIB–DNA module and the mobile regions of Pol II (clamp and stalk) are in different relative positions. Whereas the TBP–TFIIA–TFIIB module and the clamp and stalk element resembles those in the 'synthetic' model (**a–c**), the path of the DNA is very different, in that it moves away, rather than towards the Pol II stalk.
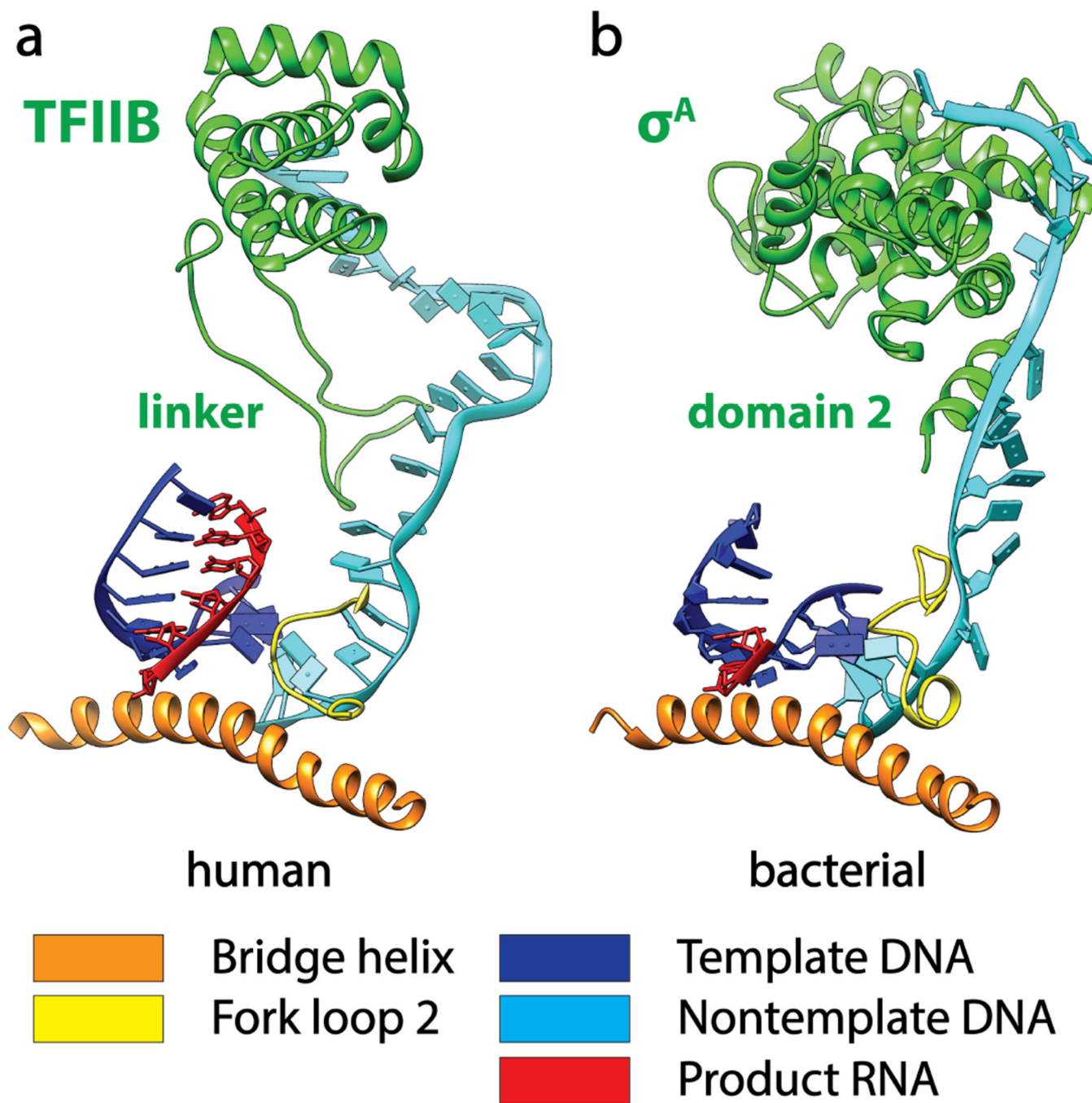
Extended Data Figure 8. Comparison of eukaryotic and bacterial initiation complexes around the active site

a, **b**, Equivalent close-up views of our ITC(–IIS) model (**a**) and the crystallographic structure of a bacterial initiation complex (4G7O[42]) (**b**). The fork loop 2 is tilted in a very

similar manner to that observed in our OC, ITC and ITC(–IIS) structures (see Figs 2b and 5f). Domain 2 within the bacterial σ factor is involved in stabilizing the non-template DNA in a similar manner as the TFIIB linker region in human initiation complex.
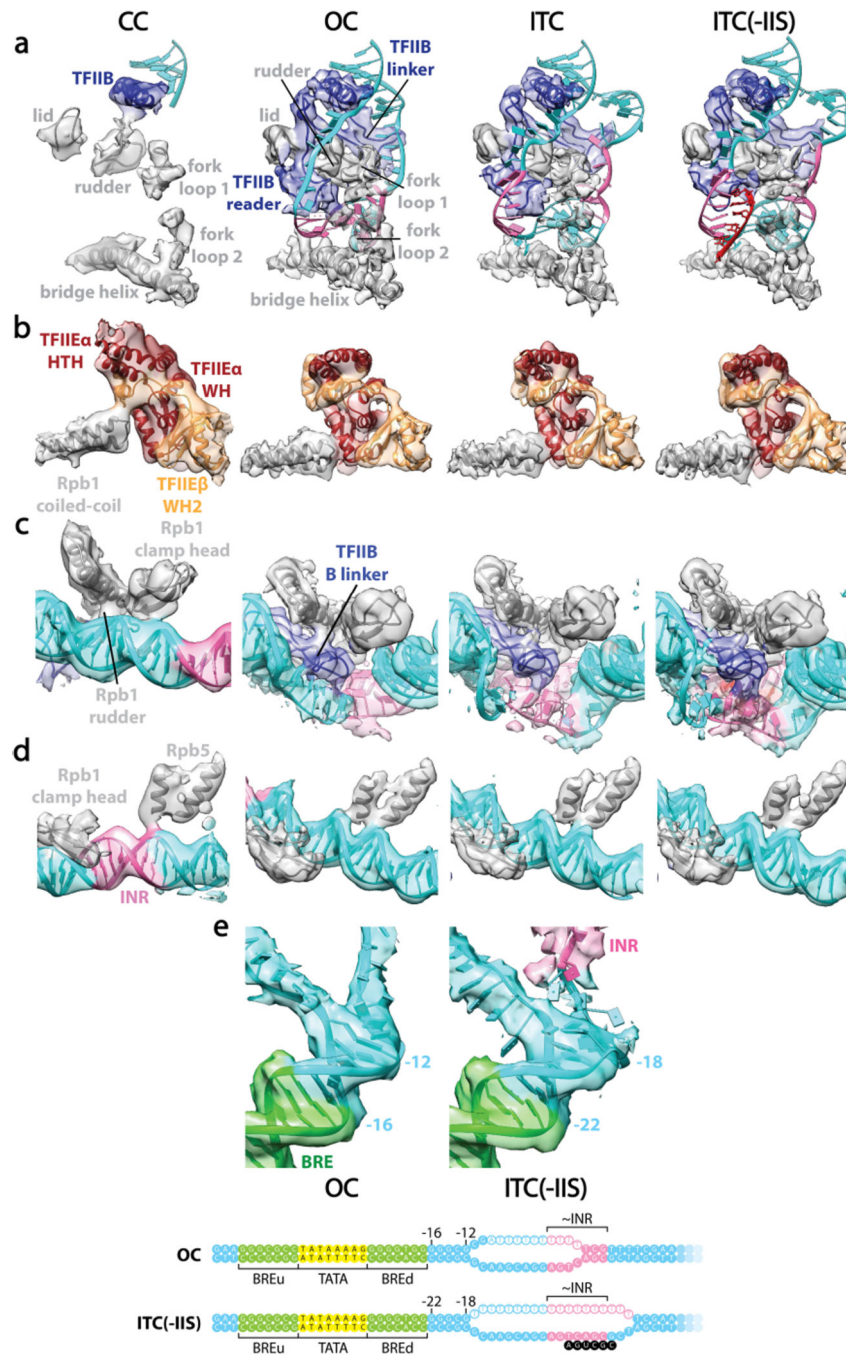
Extended Data Figure 9. Comparison of selected regions of the human core PIC structures for the CC, OC, ITC and ITC(–IIS) states

a, Segmented density and model for TFIIB and Pol II loops critical for stabilization of the transcription bubble. The density for the nucleic acids has been omitted for clarity. **b**, Interaction of the Pol II clamp head with TFIIE. As the clamp closes down during promoter

opening (CC to OC transition), the region of contact with TFIIE changes. **c**, Interaction of the Pol II clamp head with DNA near the promoter melting site. **d**, Interaction of RPB5 with DNA. **e**, Opening of two extra base pairs of DNA in the OC scaffold. EM density and the corresponding model near the initiation bubble upstream fork in the OC (left) and ITC(–IIS) (right) structures. Positions of the duplexed DNA downstream of the BRE were labelled relative to the +1 active site in each structure. The cartoon (bottom) shows the two aligned DNA templates for reference.

Extended Data Table 1

Statistics after refinement of atomic coordinates with Phenix for the core PIC models (OC, ITC, ITC(–IIS) and CC), and scores from MolProbilty structural analysis

	OC	ITC	ITC without TFIIS	CC
R	0.31	0.30	0.305	0.32
R-free	0.32	0.31	0.306	0.34
RMS angles (°)	1.41	1.49	1.48	1.10
RMS bonds (Å)	0.010	0.012	0.011	0.005
Ramachandran outliers (%)	2.8	3.0	2.67	2.6
Ramachandran favored (%)	88.4	88.7	88.55	89.9
Clashscore	12.77	12.87	14.10	22.04
Overall MolProbity score	2.42	2.50	2.44	2.54

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Roeder RG. The role of general initiation factors in transcription by RNA polymerase II. Trends Biochem. Sci. 1996; 21:327–335. [PubMed: 8870495]

2. Goodrich JA, Cutler G, Tjian R. Contacts in context: promoter specificity and macromolecular interactions in transcription. Cell. 1996; 84:825–830. [PubMed: 8601306]

3. Boeger H, et al. Structural basis of eukaryotic gene transcription. FEBS Lett. 2005; 579:899–903. [PubMed: 15680971]

4. Buratowski S, Hahn S, Guarente L, Sharp PA. Five intermediate complexes in transcription initiation by RNA polymerase II. Cell. 1989; 56:549–561. [PubMed: 2917366]

5. Cheung ACM, Sainsbury S, Cramer P. Structural basis of initial RNA polymerase II transcription. EMBO J. 2011; 30:4755–4763. [PubMed: 22056778]

6. Liu X, Bushnell DA, Silva DA, Huang X, Kornberg RD. Initiation complex structure and promoter proofreading. Science. 2011; 333:633–637. [PubMed: 21798951]

7. Luse DS, Jacob GA. Abortive initiation by RNA polymerase II *in vitro* at the adenovirus 2 major late promoter. J. Biol. Chem. 1987; 262:14990–14997. [PubMed: 3667620]

8. Liu X, Bushnell DA, Wang D, Calero G, Kornberg RD. Structure of an RNA polymerase II-TFIIB complex and the transcription initiation mechanism. Science. 2010; 327:206–209. [PubMed: 19965383]

9. 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. 2012; 19:788–796. [PubMed: 22751016]

10. Sainsbury S, Niesser J, Cramer P. Structure and function of the initially transcribing RNA polymerase II–TFIIB complex. Nature. 2013; 493:437–440. [PubMed: 23151482]

11. He Y, Fang J, Taatjes DJ, Nogales E. Structural visualization of key steps in human transcription initiation. Nature. 2013; 495:481–486. [PubMed: 23446344]

12. Plaschka C, et al. Architecture of the RNA polymerase II–Mediator core initiation complex. Nature. 2015; 518:376–380. [PubMed: 25652824]

13. Murakami K, et al. Structure of an RNA polymerase II preinitiation complex. Proc. Natl Acad. Sci. USA. 2015; 112:13543–13548. [PubMed: 26483468]

14. Louder RK, et al. Structure of promoter-bound TFIID and model of human pre-initiation complex assembly. Nature. 2016; 531:604–609. [PubMed: 27007846]

15. Kim B, et al. The transcription elongation factor TFIIS is a component of RNA polymerase II preinitiation complexes. Proc. Natl Acad. Sci. USA. 2007; 104:16068–16073. [PubMed: 17913884]

16. Wind M, Reines D. Transcription elongation factor SII. BioEssays. 2000; 22:327–336. [PubMed: 10723030]

17. Fish RN, Kane CM. Promoting elongation with transcript cleavage stimulatory factors. Biochim. Biophys. Acta. 2002; 1577:287–307. [PubMed: 12213659]

18. Kostrewa D, et al. RNA polymerase II-TFIIB structure and mechanism of transcription initiation. Nature. 2009; 462:323–330. [PubMed: 19820686]

19. Buratowski S, Zhou H. Transcription factor IID mutants defective for interaction with transcription factor IIA. Science. 1992; 255:1130–1132. [PubMed: 1546314]

20. Lee DK, DeJong J, Hashimoto S, Horikoshi M, Roeder RG. TFIIA induces conformational changes in TFIID via interactions with the basic repeat. Mol. Cell. Biol. 1992; 12:5189–5196. [PubMed: 1406690]

21. Bleichenbacher M, Tan S, Richmond TJ. Novel interactions between the components of human and yeast TFIIA/TBP/DNA complexes. J. Mol. Biol. 2003; 332:783–793. [PubMed: 12972251]

22. Juo ZS, Kassavetis GA, Wang J, Geiduschek EP, Sigler PB. Crystal structure of a transcription factor IIIB core interface ternary complex. Nature. 2003; 422:534–539. [PubMed: 12660736]

23. Wollmann P, et al. Structure and mechanism of the Swi2/Snf2 remodeller Mot1 in complex with its substrate TBP. Nature. 2011; 475:403–407. [PubMed: 21734658]

24. Anandapadamanaban M, et al. High-resolution structure of TBP with TAF1 reveals anchoring patterns in transcriptional regulation. Nat. Struct. Mol. Biol. 2013; 20:1008–1014. [PubMed: 23851461]

25. Wang D, Bushnell DA, Westover KD, Kaplan CD, Kornberg RD. Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. Cell. 2006; 127:941–954. [PubMed: 17129781]

26. Kettenberger H, Armache KJ, Cramer P. Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. Mol. Cell. 2004; 16:955–965. [PubMed: 15610738]

27. Chen ZA, et al. Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry. EMBO J. 2010; 29:717–726. [PubMed: 20094031]

28. Eichner J, Chen HT, Warfield L, Hahn S. Position of the general transcription factor TFIIF within the RNA polymerase II transcription preinitiation complex. EMBO J. 2010; 29:706–716. [PubMed: 20033062]

29. Ohkuma Y, Hashimoto S, Wang CK, Horikoshi M, Roeder RG. Analysis of the role of TFIIE in basal transcription and TFIIH-mediated carboxy-terminal domain phosphorylation through structure-function studies of TFIIE-alpha. Mol. Cell. Biol. 1995; 15:4856–4866. [PubMed: 7651404]

30. Schultz P, et al. Molecular structure of human TFIIH. Cell. 2000; 102:599–607. [PubMed: 11007478]

31. Gibbons BJ, et al. Subunit architecture of general transcription factor TFIIH. Proc. Natl Acad. Sci. USA. 2012; 109:1949–1954. [PubMed: 22308316]

32. Luo J, et al. Architecture of the human and yeast general transcription and DNA repair factor TFIIH. Mol. Cell. 2015; 59:794–806. [PubMed: 26340423]

33. Dürr H, Körner C, Müller M, Hickmann V, Hopfner KP. X-ray structures of the Sulfolobus solfataricus SWI2/SNF2 ATPase core and its complex with DNA. Cell. 2005; 121:363–373. [PubMed: 15882619]

34. Lewis R, Dürr H, Hopfner KP, Michaelis J. Conformational changes of a Swi2/Snf2 ATPase during its mechano-chemical cycle. Nucleic Acids Res. 2008; 36:1881–1890. [PubMed: 18267970]

35. Thomä NH, et al. Structure of the SWI2/SNF2 chromatin-remodeling domain of eukaryotic Rad54. Nat. Struct. Mol. Biol. 2005; 12:350–356. [PubMed: 15806108]

36. Coin F, et al. p8/TTD-A as a repair-specific TFIIH subunit. Mol. Cell. 2006; 21:215–226. [PubMed: 16427011]

37. Coin F, Oksenych V, Egly JM. Distinct roles for the XPB/p52 and XPD/p44 subcomplexes of TFIIH in damaged DNA opening during nucleotide excision repair. Mol. Cell. 2007; 26:245–256. [PubMed: 17466626]

38. Tsai FTF, Sigler PB. Structural basis of preinitiation complex assembly on human pol II promoters. EMBO J. 2000; 19:25–36. [PubMed: 10619841]

39. Gnatt AL, Cramer P, Fu J, Bushnell DA, Kornberg RD. Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. Science. 2001; 292:1876–1882. [PubMed: 11313499]

40. Buratowski S, Sopta M, Greenblatt J, Sharp PA. RNA polymerase II-associated proteins are required for a DNA conformation change in the transcription initiation complex. Proc. Natl Acad. Sci. USA. 1991; 88:7509–7513. [PubMed: 1881889]

41. Struhl K. Promoters, activator proteins, and the mechanism of transcriptional initiation in yeast. Cell. 1987; 49:295–297. [PubMed: 2882858]

42. Zhang Y, et al. Structural basis of transcription initiation. Science. 2012; 338:1076–1080. [PubMed: 23086998]

43. Giardina C, Lis JT. DNA melting on yeast RNA polymerase II promoters. Science. 1993; 261:759–762. [PubMed: 8342041]

44. Kim TK, Ebright RH, Reinberg D. Mechanism of ATP-dependent promoter melting by transcription factor IIH. Science. 2000; 288:1418–1421. [PubMed: 10827951]

45. Bernecky C, Herzog F, Baumeister W, Plitzko JM, Cramer P. Structure of transcribing mammalian RNA polymerase II. Nature. 2016; 529:551–554. [PubMed: 26789250]

46. abart P, Újvári A, Pal M, Luse DS. Transcription factor TFIIF is not required for initiation by RNA polymerase II, but it is essential to stabilize transcription factor TFIIB in early elongation complexes. Proc. Natl Acad. Sci. USA. 2011; 108:15786–15791. [PubMed: 21896726]

47. Kapanidis AN, et al. Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. Science. 2006; 314:1144–1147. [PubMed: 17110578]

48. Revyakin A, Liu C, Ebright RH, Strick TR. Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. Science. 2006; 314:1139–1143. [PubMed: 17110577]

49. Iyer LM, Koonin EV, Aravind L. Evolutionary connection between the catalytic subunits of DNA-dependent RNA polymerases and eukaryotic RNA-dependent RNA polymerases and the origin of RNA polymerases. BMC Struct. Biol. 2003; 3:1. [PubMed: 12553882]

50. Revyakin A, et al. Transcription initiation by human RNA polymerase II visualized at single-molecule resolution. Genes Dev. 2012; 26:1691–1702. [PubMed: 22810624]

51. Kassube SA, Jinek M, Fang J, Tsutakawa S, Nogales E. Structural mimicry in transcription regulation of human RNA polymerase II by the DNA helicase RECQL5. Nat. Struct. Mol. Biol. 2013; 20:892–899. [PubMed: 23748380]

52. Juven-Gershon T, Cheng S, Kadonaga JT. Rational design of a super core promoter that enhances gene expression. Nat. Methods. 2006; 3:917–922. [PubMed: 17124735]

53. Pal M, Ponticelli AS, Luse DS. The role of the transcription bubble and TFIIB in promoter clearance by RNA polymerase II. Mol. Cell. 2005; 19:101–110. [PubMed: 15989968]

54. Suloway C, et al. Automated molecular microscopy: the new Leginon system. J. Struct. Biol. 2005; 151:41–60. [PubMed: 15890530]

55. Li X, et al. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. Nat. Methods. 2013; 10:584–590. [PubMed: 23644547]

56. Lander GC, et al. Appion: an integrated, database-driven pipeline to facilitate EM image processing. J. Struct. Biol. 2009; 166:95–102. [PubMed: 19263523]

57. Voss NR, Yoshioka CK, Radermacher M, Potter CS, Carragher B. DoG Picker and TiltPicker: software tools to facilitate particle selection in single particle electron microscopy. J. Struct. Biol. 2009; 166:205–213. [PubMed: 19374019]

58. Mindell JA, Grigorieff N. Accurate determination of local defocus and specimen tilt in electron microscopy. J. Struct. Biol. 2003; 142:334–347. [PubMed: 12781660]

59. Scheres SH. RELION: implementation of a Bayesian approach to cryo-EM structure determination. J. Struct. Biol. 2012; 180:519–530. [PubMed: 23000701]

60. Henderson R, et al. Outcome of the first electron microscopy validation task force meeting. Structure. 2012; 20:205–214. [PubMed: 22325770]

61. Scheres SH. Beam-induced motion correction for sub-megadalton cryo-EM particles. eLife. 2014; 3:e03665. [PubMed: 25122622]

62. Heymann JB, Belnap DM. Bsoft: image processing and molecular modeling for electron microscopy. J. Struct. Biol. 2007; 157:3–18. [PubMed: 17011211]

63. Cardone G, Heymann JB, Steven AC. One number does not fit all: mapping local variations in resolution in cryo-EM reconstructions. J. Struct. Biol. 2013; 184:226–236. [PubMed: 23954653]

64. Goddard TD, Huang CC, Ferrin TE. Visualizing density maps with UCSF Chimera. J. Struct. Biol. 2007; 157:281–287. [PubMed: 16963278]

65. Sali A. Comparative protein modeling by satisfaction of spatial restraints. Mol. Med. Today. 1995; 1:270–277. [PubMed: 9415161]

66. Martí-Renom MA, et al. Comparative protein structure modeling of genes and genomes. Annu. Rev. Biophys. Biomol. Struct. 2000; 29:291–325. [PubMed: 10940251]

67. Meka H, Werner F, Cordell SC, Onesti S, Brick P. Crystal structure and RNA binding of the Rpb4/Rpb7 subunits of human RNA polymerase II. Nucleic Acids Res. 2005; 33:6435–6444. [PubMed: 16282592]

68. Gaiser F, Tan S, Richmond TJ. Novel dimerization fold of RAP30/RAP74 in human TFIIF at 1.7 A resolution. J. Mol. Biol. 2000; 302:1119–1127. [PubMed: 11183778]

69. Adamczak R, Porollo A, Meller J. Accurate prediction of solvent accessibility using neural networks-based regression. Proteins. 2004; 56:753–767. [PubMed: 15281128]

70. Buchan DWA, Minneci F, Nugent TCO, Bryson K, Jones DT. Scalable web services for the PSIPRED Protein Analysis Workbench. Nucleic Acids Res. 2013; 41:W349–W357. [PubMed: 23748958]

71. Groft CM, Uljon SN, Wang R, Werner MH. Structural homology between the Rap30 DNA-binding domain and linker histone H5: implications for preinitiation complex assembly. Proc. Natl Acad. Sci. USA. 1998; 95:9117–9122. [PubMed: 9689043]

72. Meinhart A, Blobel J, Cramer P. An extended winged helix domain in general transcription factor E/IIE alpha. J. Biol. Chem. 2003; 278:48267–48274. [PubMed: 13679366]

73. Shinkai A, et al. The putative DNA-binding protein Sto12a from the thermoacidophilic archaeon Sulfolobus tokodaii contains intrachain and interchain disulfide bonds. J. Mol. Biol. 2007; 372:1293–1304. [PubMed: 17720190]

74. Okuda M, et al. A novel zinc finger structure in the large subunit of human general transcription factor TFIIE. J. Biol. Chem. 2004; 279:51395–51403. [PubMed: 15385556]

75. Okuda M, et al. Structure of the central core domain of TFIIEbeta with a novel double-stranded DNA-binding surface. EMBO J. 2000; 19:1346–1356. [PubMed: 10716934]

76. Qian X, et al. Novel zinc finger motif in the basal transcriptional machinery: three-dimensional NMR studies of the nucleic acid binding domain of transcriptional elongation factor TFIIS. Biochemistry. 1993; 32:9944–9959. [PubMed: 8399164]

77. Fan L, et al. Conserved XPB core structure and motifs for DNA unwinding: implications for pathway selection of transcription or excision repair. Mol. Cell. 2006; 22:27–37. [PubMed: 16600867]

78. Hilario E, Li Y, Nobumori Y, Liu X, Fan L. Structure of the C-terminal half of human XPB helicase and the impact of the disease-causing mutation XP11BE. Acta Crystallogr. D. 2013; 69:237–246. [PubMed: 23385459]

79. Wolski SC, et al. Crystal structure of the FeS cluster-containing nucleotide excision repair helicase XPD. PLoS Biol. 2008; 6:e149. [PubMed: 18578568]

80. Fan L, et al. XPD helicase structures and activities: insights into the cancer and aging phenotypes from XPD mutations. Cell. 2008; 133:789–800. [PubMed: 18510924]

81. Kim JS, et al. Crystal structure of the Rad3/XPD regulatory domain of Ssl1/p44. J. Biol. Chem. 2015; 290:8321–8330. [PubMed: 25681444]

82. Kainov DE, Vitorino M, Cavarelli J, Poterszman A, Egly JM. Structural basis for group A trichothiodystrophy. Nat. Struct. Mol. Biol. 2008; 15:980–984. [PubMed: 19172752]

83. Maertens GN, et al. Structural basis for nuclear import of splicing factors by human Transportin 3. Proc. Natl Acad. Sci. USA. 2014; 111:2728–2733. [PubMed: 24449914]

84. Pei J, Kim BH, Tang M, Grishin NV. PROMALS web server for accurate multiple protein sequence alignments. Nucleic Acids Res. 2007; 35:W649–W652. [PubMed: 17452345]

85. Schmitt DR, Kuper J, Elias A, Kisker C. The structure of the TFIIH p34 subunit reveals a von Willebrand factor A like fold. PLoS One. 2014; 9:e102389. [PubMed: 25013903]

86. Lu XJ, Olson WK. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. Nucleic Acids Res. 2003; 31:5108–5121. [PubMed: 12930962]

87. Trabuco LG, Villa E, Mitra K, Frank J, Schulten K. Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. Structure. 2008; 16:673–683. [PubMed: 18462672]

88. Cornell WD, et al. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J. Am. Chem. Soc. 1995; 117:5179–5197.

89. Kalé L, et al. NAMD2: Greater scalability for parallel molecular dynamics. J. Comput. Phys. 1999; 151:283–312.

90. Hornak V, et al. Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins. 2006; 65:712–725. [PubMed: 16981200]

91. Adams PD, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D. 2010; 66:213–221. [PubMed: 20124702]

92. Adams PD, et al. The Phenix software for automated determination of macromolecular structures. Methods. 2011; 55:94–106. [PubMed: 21821126]

93. Chen VB, et al. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. D. 2010; 66:12–21. [PubMed: 20057044]

94. Okuda M, et al. Structural insight into the TFIIE-TFIIH interaction: TFIIE and p53 share the binding region on TFIIH. EMBO J. 2008; 27:1161–1171. [PubMed: 18354501]
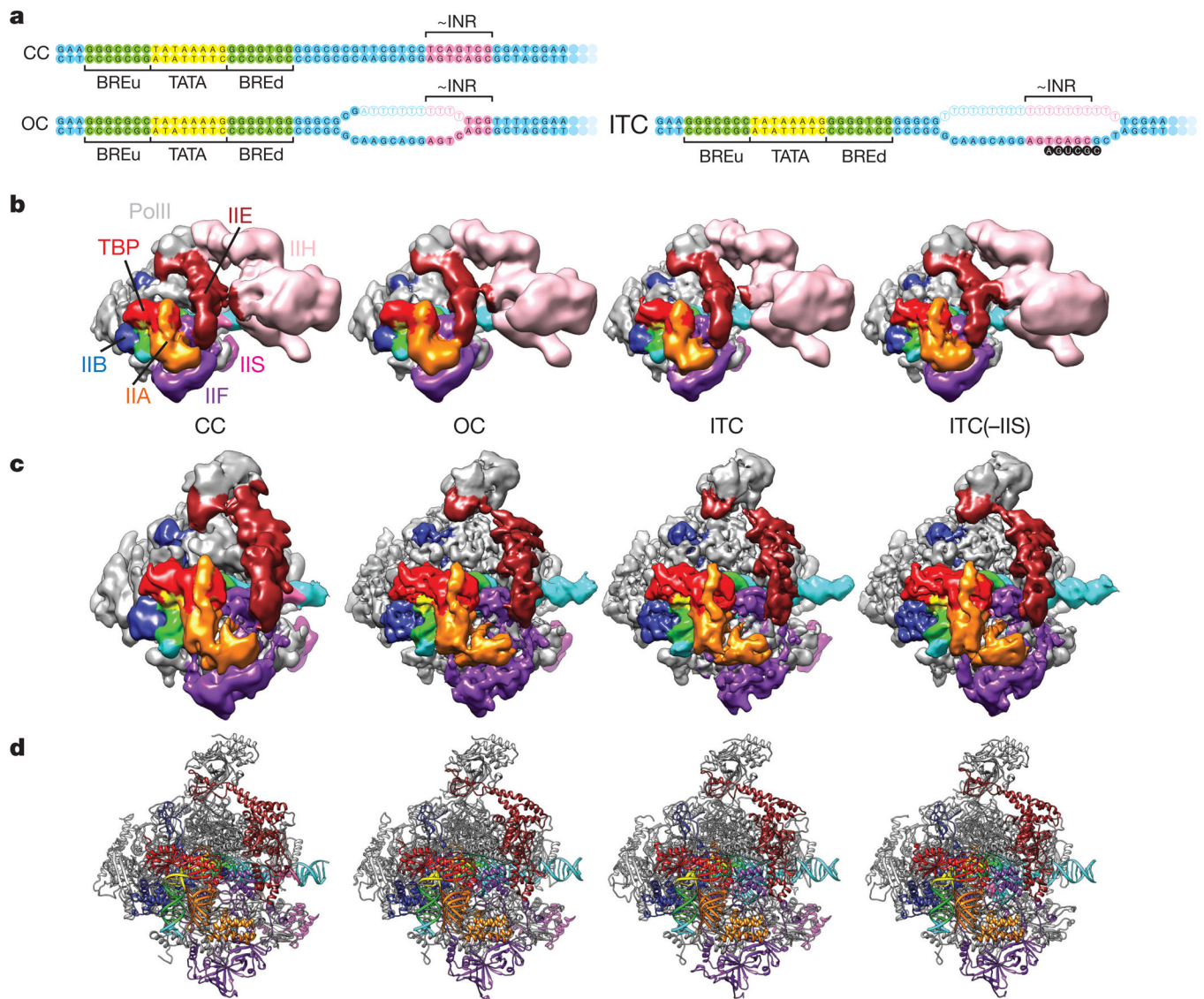
Figure 1. Cryo-EM reconstructions of human PIC in different states during the initiation process
a, Nucleic acid scaffolds used. Filled and open circles correspond, respectively, to the core promoter and poly-T mismatch sequences. For the OC, two additional bases upstream are opened in the cryo-EM structure. Black circles correspond to RNA. **b–d**, Cryo-EM reconstructions of holo-PICs (**b**), core PICs (**c**), and MDFF models (**d**), for the CC, OC, ITC and ITC(–IIS) states.
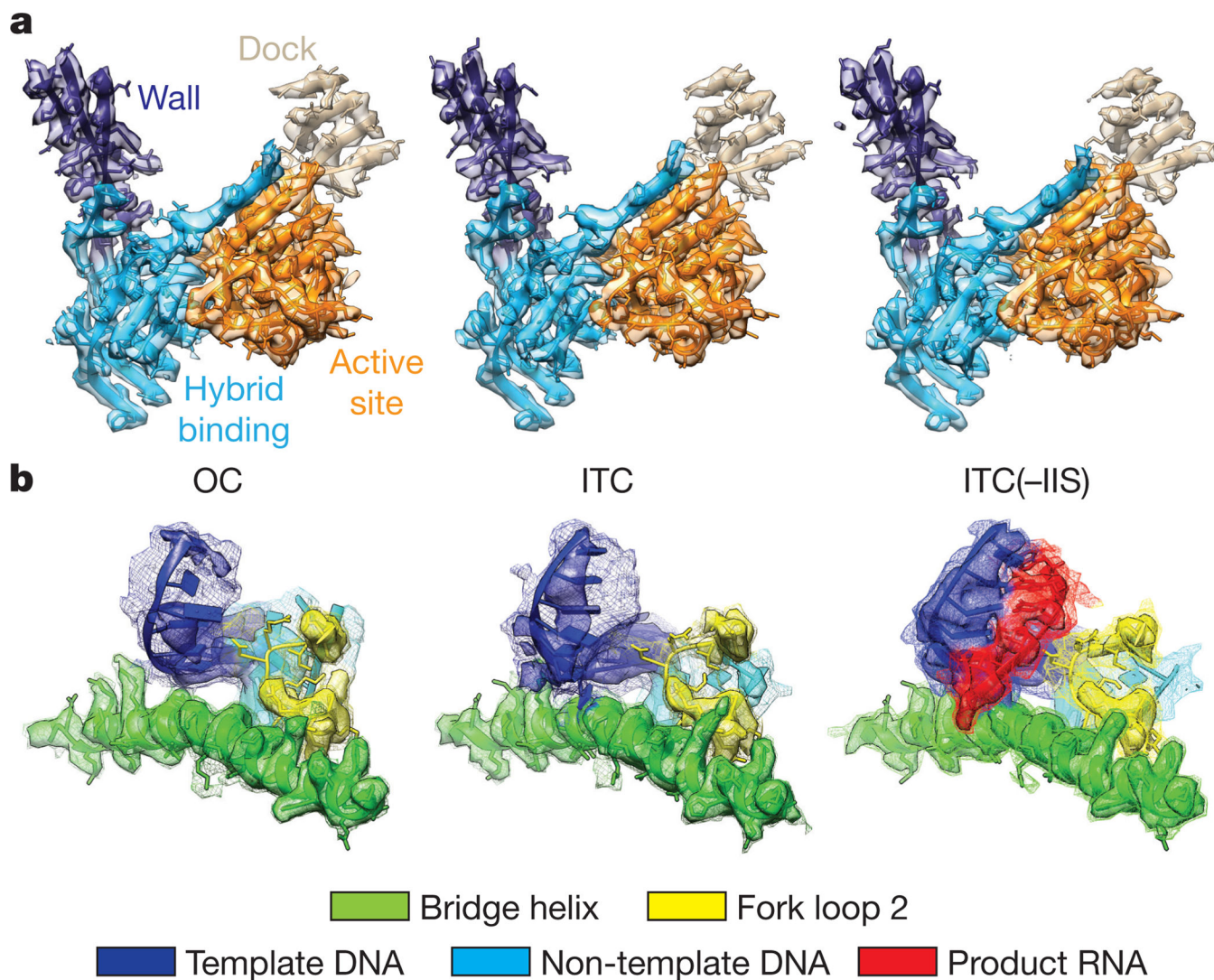
Figure 2. Examples of near-atomic resolution regions

a, EM densities and corresponding atomic models for the double-ψ β-barrel domain composing the conserved core of Pol II within RPB1 and RPB2 (ref. 49) (see also Supplementary Video 1). **b**, Structural details around the Pol II active site. Densities are shown at two different thresholds. The lower threshold (mesh) allows visualization of the more flexible elements.
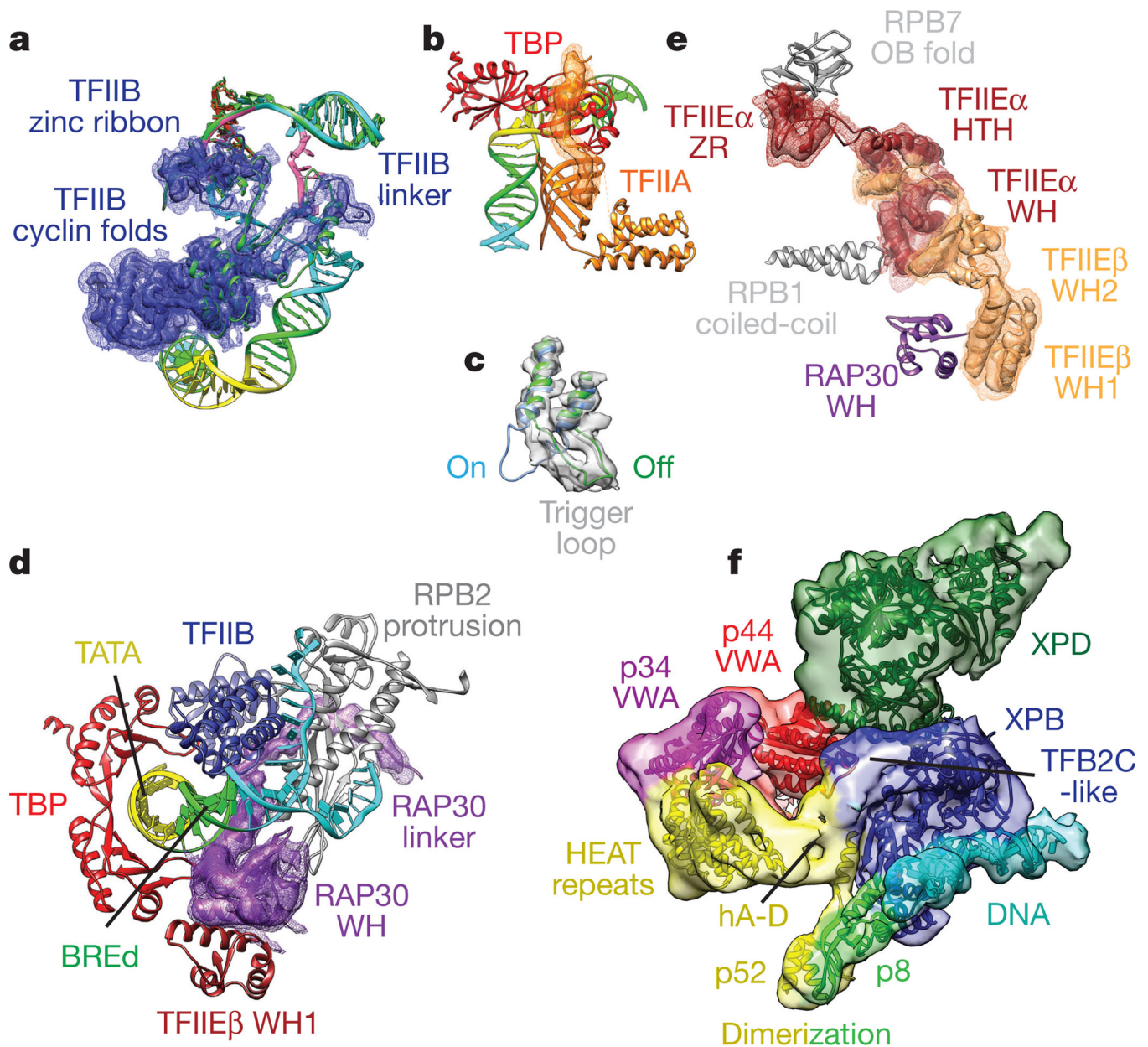
Figure 3. Newly visualized structural elements of the human PIC
Segmented cryo-EM densities from the ITC(−IIS) reconstruction shown at two different thresholds (the lower threshold (mesh) facilitates visualization of the more flexible elements). **a**, Differences in the path of the B-linker of TFIIB between the yeast crystallographic model[10] (green) and the human ITC cryo-EM structure (coloured as in the rest of the figures). Superimposition was performed by aligning the rigid part of Pol II (that is, excluding the clamp and the stalk), here and for all other figures. **b**, Residues 307–332 of TFIIAβ interact with the N-terminal lobe of TBP. **c**, Trigger loop in RPB1 in an open 'off' state. The trigger loops in crystal structures of yeast EC in its closed 'on' state (2NVZ[25], light blue), and in its open 'off' state (1Y1V[26], green) are shown for comparison. **d**, RAP30 WH domain and linker within TFIIF. Elements of the PIC in close proximity are indicated.

e, Key domains of TFIIE and their interaction partners within the PIC. **f**, Cryo-EM structure of TFIIH (the flexible CAK subcomplex is not visible after averaging) obtained from a combined data set of open promoter states. Segmentation based on fitting of available structures and homology models.
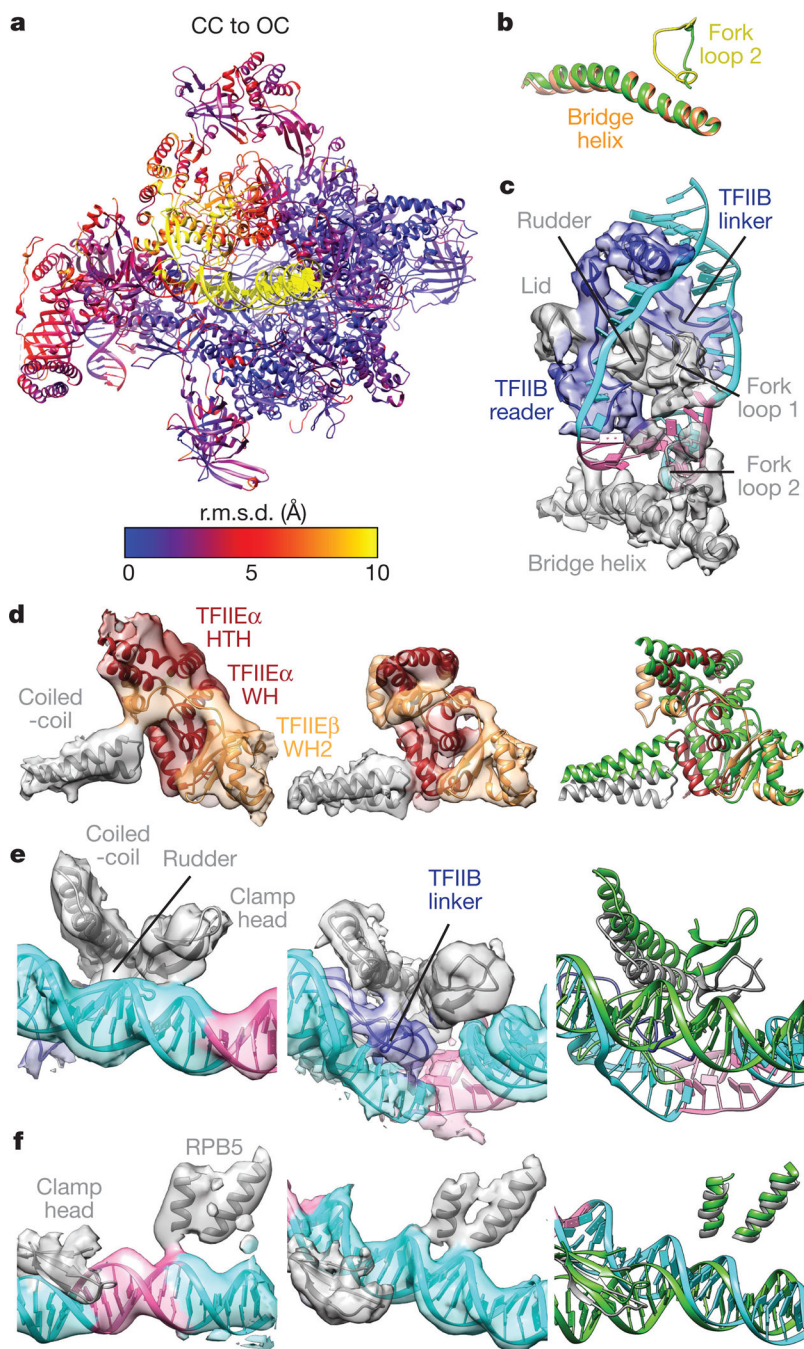
Figure 4. Structural transitions during promoter opening

a, Colour-coded motion of backbone atoms between CC and OC based on r.m.s.d. **b**, Change of fork loop 2 near the active site between the CC (green) and OC (coloured) (see also Fig. 2b). A similar colour scheme is used for the superposition of CC and OC states in the third column of **d–f**. **c**, Segmented cryo-EM density from the OC reconstruction and the corresponding model for TFIIB and various Pol II loops critical for stabilization of the transcription bubble. The density for the nucleic acids has been omitted for clarity. **d**, Changes in the contact between the RPB1 coiled-coil and TFIIE during bubble opening. **e**,

Changes near the melting start site involving the clamp head, coiled-coil domain, rudder domain and TFIIB. **f**, Changes in Pol II interaction with downstream DNA involving the clamp head and RPB5.
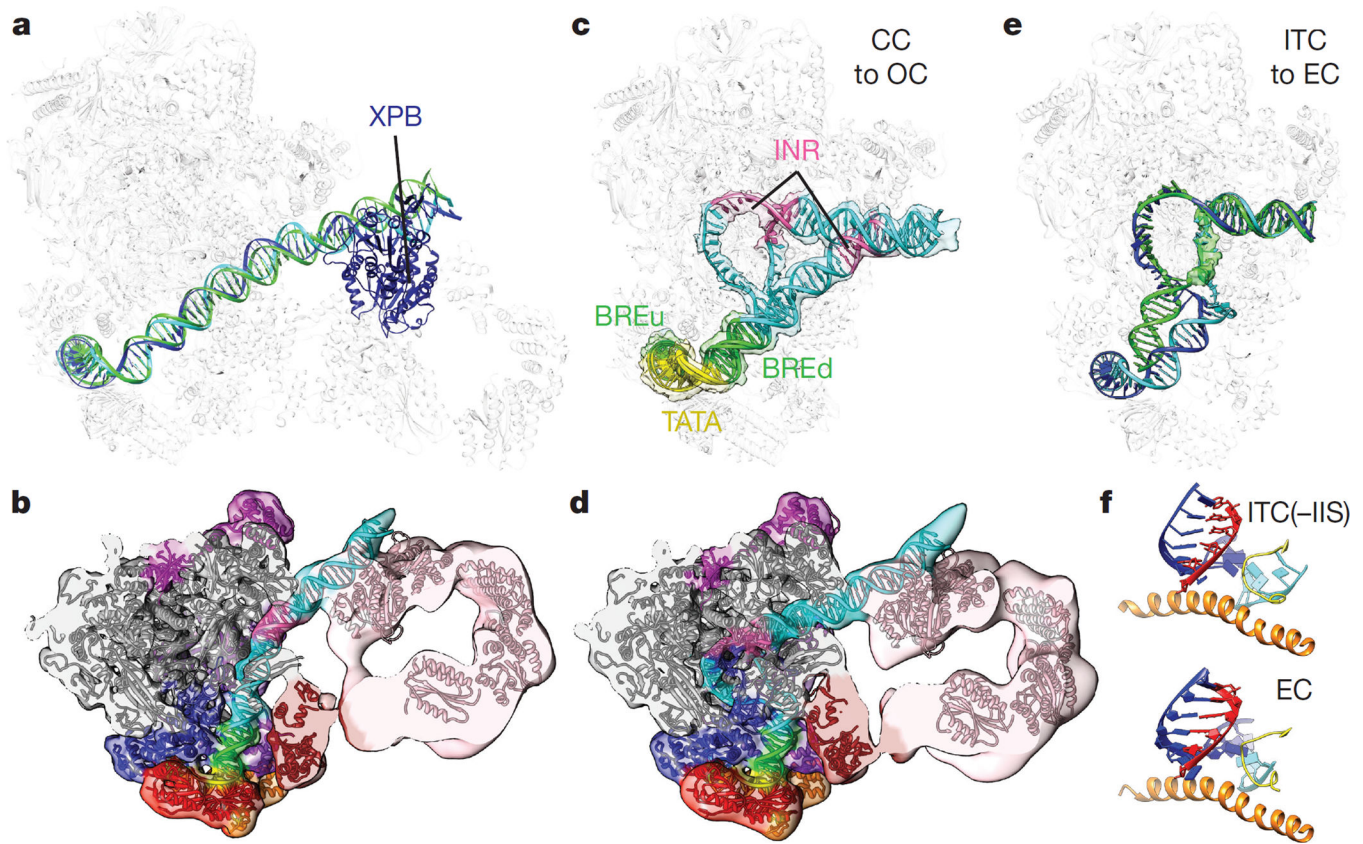
Figure 5. Nucleic acids rearrangements between stages of transcription

a, DNA paths before[11] (green) and after (blue/cyan) TFIIH incorporation into the CC. XPB is coloured navy blue and the rest of the PIC is shown in transparency. **b**, Section through the CC density map and model showing the path of the DNA and its engagement by TFIIH. **c**, Segmented electron microscopy densities and models of promoter DNA in the CC and OC. **d**, Section through the OC density map and model showing the path of the DNA and its engagement by TFIIH. **e**, Comparison of the DNA path in our ITC(−IIS) state (blue/cyan) with that in the bovine EC structure[45] (green). Segmented electron microscopy density of the EC single-stranded nontemplate DNA is shown to illustrate its path, as this part was not modelled in the published structure. **f**, Close-up views at the active site comparing the position of fork loop 2 relative to the nucleic acids and bridge helix in the human ITC(−IIS) (top) and bovine EC (bottom).
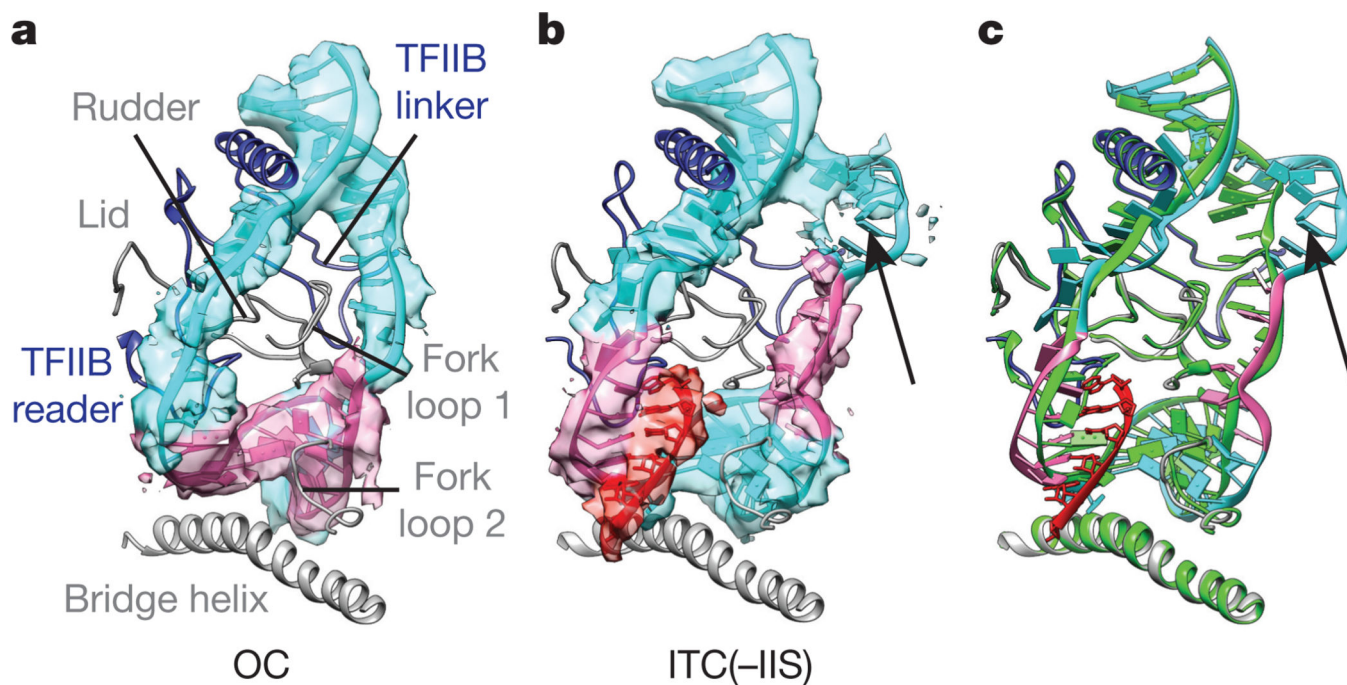
Figure 6. Possible region of DNA scrunching during initiation
a, b, Segmented cryo-EM densities and models for DNA and RNA near the Pol II active site for the OC (**a**) and ITC(−IIS) (**b**). TFIIB and Pol II loops that are critical for stabilization of the bubble are also shown (density omitted for clarity). **c**, Superimposition of the OC (green) and ITC(−IIS) (coloured) states. The lack of visible density (arrow) is interpreted as disorder at the site of scrunching, upstream of the contact with the TFIIB linker.