HHS Public Access

Author manuscript

Nature. Author manuscript; available in PMC 2013 September 28.

Published in final edited form as:

Nature. 2013 March 28; 495(7442): 481–486. doi:10.1038/nature11991.

Structural visualization of key steps in human transcription initiation

Yuan He[1], **Jie Fang**[2], **Dylan J. Taatjes**[3], and **Eva Nogales**[1,2,4,5]

[1]Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720

[2]Howard Hughes Medical Institute, University of California, Berkeley, CA 94720

[3]Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80303

[4]QB3 Institute and Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720

Abstract

Eukaryotic transcription initiation requires the assembly of general transcription factors into a pre-initiation complex that ensures the accurate loading of RNA polymerase II at the transcription start site. The molecular mechanism and function of this assembly have remained elusive due to lack of structural information. We have used an *in vitro* reconstituted system to study the stepwise assembly of human TBP, TFIIA, TFIIB, Pol II, TFIIF, TFIIE, and TFIIH onto promoter DNA using cryo-electron microscopy. Our structural analyses provide pseudo-atomic models at various stages of transcription initiation that illuminate critical molecular interactions, including how TFIIF engages Pol II and promoter DNA to stabilize both the closed PIC and the open-promoter complex and regulate start site selection. Comparison of open versus closed pre-initiation complexes, combined with the localization of the TFIIH helicases XPD and XPB, supports a DNA translocation model of XPB and explains its essential role in promoter opening.

Accurate and regulated initiation of eukaryotic gene transcription represents a major step in gene regulation, requiring the coordinated activity of a large number of proteins and protein complexes. The basal transcriptional machinery includes RNA polymerase II (Pol II) along with a series of general transcription factors (GTFs) (TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH) that assemble into a ~2 MDa complex on core promoter DNA. This pre-initiation complex (PIC) is essential to direct accurate transcription start site (TSS) selection, promoter melting, and Pol II promoter escape[1–3]. Despite recent structural advances on Pol

[5]Correspondence: ENogales@lbl.gov.

Author Contributions Y.H. designed and carried out the experiments, J.F. and D.J.T. provided essential reagents, Y.H. and E.N. analyzed the data and wrote the paper.

Author Information Cryo-EM density maps have been deposited in the Electron Microscopy Data Bank (EMDB) under accession numbers EMD-2304 (TBP-TFIIA-TFIIB-DNA-Pol II), EMD-2305 (TBP-TFIIA-TFIIB-DNA-Pol II-TFIIF), EMD-2306 (TBP-TFIIA-TFIIB-DNA-Pol II-TFIIF-TFIIE), and EMD-2307 (TBP-TFIIA-TFIIB-DNA-Pol II-TFIIF-TFIIE in the OC mimic state). Negative stain EM density maps have been assigned accession numbers EMD-2308 (TBP-TFIIA-TFIIB-DNA-Pol II-TFIIF-TFIIE-TFIIH), and EMD-2309 (apo TFIIH).

II[4,5] and subcomplexes of the PIC[6], the molecular assembly details of this essential complex remain elusive.

In vitro reconstitution of this process has provided a model for the sequential assembly pathway of transcription initiation. TFIID is the first factor specifically recruited to the promoter. This megadalton complex includes the TATA binding protein (TBP), which is sufficient for basal transcription on TATA box containing promoters[2,3,7]. TFIIA and TFIIB are then recruited, further stabilizing the interaction between TBP and promoter DNA. Next Pol II, likely in association with TFIIF, adds to the growing PIC. Finally TFIIE and TFIIH, which is required for DNA melting, are recruited to form the transcriptionally competent PIC[2,3].

Structural characterization of PIC assemblies is challenging and has been limited to a small number of electron microscopy (EM) studies[8–10]. Crystallographic structures of individual components, combined with biochemical data, have led to a number of structural models for PIC subcomplexes, in either a closed or open-promoter conformation[6,9,11,12]. In spite of this progress, important questions remain unanswered, such as how TFIIB and TFIIF serve complementary roles during the promoter opening process or how TFIIE positions TFIIH in a configuration capable of melting the DNA.

Here we present cryo-EM snapshots of PIC intermediates during sequential assembly. A reconstitution system allowed us to localize each GTF within the cryo-EM structures, track the effect of each additional factor on the PIC, and ultimately reveal the network of protein-protein and protein-DNA interactions governing PIC assembly. Furthermore, by visualization of an open-promoter complex (OC) mimic, we have obtained new mechanistic details concerning promoter melting. Altogether, our structures provide unprecedented insights into the molecular assembly, organization, and functional roles of different GTFs during transcription initiation.

Stepwise assembly/visualization of human PIC

In order to structurally characterize the sequential assembly of GTFs necessary for human transcription initiation, we developed an *in vitro* system for reconstitution and purification of a simplified PIC, in which TBP substituted for TFIID, and that ultimately contained 31 polypeptides. Our promoter DNA contained TATA, BRE and INR core promoter elements and was immobilized on streptavidin beads (Fig. 1a). After stepwise assembly of PIC intermediates by sequential incubation with the desired GTFs, stable complexes were released by restriction enzyme digestion. The effectiveness of this approach for structural characterization of the PIC intermediates was initially tested by single particle EM of negatively stained samples (Supplementary Fig. 1). This initial analysis allowed us to localize each GTF within the context of the full assembly (Fig. 1b–e), although it precluded the visualization of DNA. The stepwise purification approach enabled us to describe the effect of factor addition on the rest of the PIC, which cannot be achieved by studying individual factors or the complete PIC. The negative stain structures were then used as starting references to generate cryo-EM reconstructions of the PIC subcomplexes with

improved resolution that allowed visualization of the DNA and accurate docking of existing crystal structures (Supplementary Figs 2–5).

Recruitment of Pol II onto promoter DNA

To start, we obtained the cryo-EM structure of a PIC subcomplex containing TBP, TFIIA, TFIIB, Pol II, and core promoter DNA (Fig. 2a). Crystal structures of TBP-TFIIA-DNA[13], TBP-TFIIB-DNA[14] and yeast Pol II-TFIIB[11,12,15] could be unambiguously docked into our density map as rigid bodies (Supplementary Fig. 2e). This procedure validated our cryo-EM structure while also allowing the localization of each factor to generate a pseudo-atomic model of the assembly. The visible portion of DNA accounts only for the upstream core promoter elements, which are stabilized by protein-DNA interactions, whereas the DNA downstream of the BREd lacks contact with the PIC and was not visualized due to its flexibility (Fig. 2a).

A yeast PIC model has been previously proposed based on superimposing crystal structures using the common protein as an anchor point[12]. Our data shows that a simple pivoting of the C-terminal cyclin fold domain of TFIIB around the N-terminal one can explain the position of the TBP/TFIIA module in our map using the available crystal structures, without disrupting the interaction between the N-terminal cyclin fold of TFIIB and Pol II (Supplementary Fig. 7). This small discrepancy with the previous piece-wise model is unlikely due to differences between the human and yeast systems, but rather a re-organization with respect to individual crystal structures upon interaction of GTFs on the core promoter.

Effect of TFIIF on PIC assembly

According to the sequential assembly pathway, TFIIF is recruited to the promoter in association with Pol II[2]. In order to understand its structural role during PIC assembly, we added TFIIF separately to our reconstituted system. By comparing the cryo-EM structures of PIC subcomplexes in the absence and presence of TFIIF, we identified additional protein densities appearing at two nearby locations, by the lobe and protrusion domains of Pol II (Fig. 2b). Importantly, the addition of TFIIF also resulted in the stabilization of the downstream DNA along the cleft of Pol II, in a position that is distinct from a previously proposed model[12] (Supplementary Fig. 8). Thus, TFIIF is required for the engagement of DNA by Pol II within the context of a closed PIC.

The crystal structure of the human TFIIF dimerization domain[16] could be unambiguously fitted into the new density ascribed to TFIIF by the lobe domain of Pol II using rigid-body docking, only slightly shifted from previous models that were based on crosslinking data[17,18] (Fig. 2c and Supplementary Fig. 3e). No obvious density was observed for the arm domain of RAP74, which extends about 45 Å from the end of the RAP74 barrel[16], suggesting it is mobile at this stage of PIC assembly. A small clash between the RPB2 lobe and the RAP74 α1 helix, can be explained by a reorganization of this element in the context of the PIC (see Supplementary Fig. 9 for details).

In addition to interacting with Pol II, our cryo-EM structure indicates that a region of TFIIF directly contacts the BREd (Fig. 2c,d). We propose that this region corresponds to the C-terminal winged-helix (WH) domain of RAP30, based upon the following: 1) the size of this additional TFIIF density is consistent with the RAP30 WH domain; 2) RAP30 has been shown to crosslink to BREd[19] and its C-terminal WH domain has been identified to be in direct contact with the protrusion domain of Pol II[18]; 3) RAP30, rather than RAP74, is required for accurate transcription initiation[20] and deletion mutants of RAP30's WH domain are lethal in yeast[18,21]. This WH domain therefore contributes to a unique nucleoprotein complex formed by TBP, TFIIA, and the TFIIB cyclin fold, as they contact the core promoter elements upstream of the INR, which is further stabilized by the protrusion domain of Pol II (Fig. 2d).

Comparison of structures shown in Fig. 2a and Fig. 2b strongly suggests that the overall effect of TFIIF on the assembling PIC is a clear stabilization of the DNA along the Pol II cleft. Given its position, we propose that the RAP30 WH domain plays an essential role in positioning the flexible promoter DNA downstream of BREd along the Pol II cleft, thus facilitating subsequent steps in the promoter melting process. However, a contribution to this DNA stabilization by the dimerization domain is also possible. Correct positioning of the DNA by TFIIF is consistent with its role in promoter opening and TSS selection[22–24]. Our structures also revealed the opening of the Pol II clamp domain as it accommodates the downstream DNA (Fig. 2e and Supplementary Figure 11). In addition, we observe further rotation of the TBP-TFIIA-TFIIB subcomplex with respect to previously proposed models, positioning it even closer to Pol II (Fig. 2e and Supplementary Fig. 7).

TFIIE-containing PIC and DNA contacts

The addition of TFIIE to the growing PIC resulted in new protein density that connects TFIIF with Pol II's stalk domain (Fig. 3a). The resolution of this reconstruction (11 Å) was the highest obtained for any of the complexes studied, suggesting that TFIIE stabilizes the PIC. The density corresponding to TFIIE, however, was the least well-defined element according to local resolution calculations (Supplementary Fig. 4f), which may be due to flexible connections between the WH domains predicted within the TFIIE structure[6] (Fig. 3a,b). One end of TFIIE associates with the stalk of Pol II by interacting with the RPB7 L45 loop (Fig. 3c), which has been predicted to stabilize the OC and whose deletion completely abolished transcription[25]. Also consistent with the positioning of TFIIE in our PIC structure, a zinc ribbon domain within the archaeal homolog of TFIIE was found to be located near the base of the stalk domain of the polymerase[26]. Away from the stalk, the TFIIE density contacts the Pol II clamp domain to interact ultimately with the WH domain of TFIIF. A model of the three WH domains within TFIIE interacting with elements of the clamp head has been proposed based on crosslinking studies[6] (Fig. 3c). Although the model cannot fit the EM density perfectly, the overall path of the three tandem WH domains in the model follows the elongated TFIIE cryo-EM density and ends by directly contacting the RAP30 WH domain (Fig. 3b). Therefore, a continuous chain of four WH domains appears to link the Pol II clamp region with the TBP-TFIIA-TFIIB-DNA subcomplex, preventing DNA from leaving the cleft.

Our 11 Å resolution reconstruction of the PIC containing TFIIE starts to reveal the major and minor grooves of the promoter DNA (Fig. 3d), allowing us to model its path. We found that linear B-form DNA could not be accommodated into the DNA density (Supplementary Fig. 12), requiring instead a smooth bend of 18° between positions −23 and +7 that fitted both the path and groove features of the EM density. Interestingly, a hypersensitivity region around −6 position[27] locates at one of the downstream DNA-Pol II interfaces as discussed below.

We observed two protein contacts with the downstream DNA. One connection involves the 3-strand β sheet below the clamp head while the other is mediated by a 2-helix bundle at the tip of the RPB5 jaw (Fig. 3d). Interestingly, these are the only two positively charged protein surfaces on Pol II along the path of the downstream DNA (Supplementary Fig. 12). The INR element is sandwiched precisely between these two protein-DNA contacts, an arrangement that may be relevant in promoter melting at the correct position in the DNA. The slightly open clamp conformation seen upon DNA placement onto the cleft following TFIIF addition is likely due to the interaction of the DNA with the clamp head β sheet (Fig. 3d and Supplementary Fig. 11b,c).

Whereas the spacing between the TATA box and the TSS can vary between species, the region within promoter DNA that is melted during transcription initiation is ~20 bp downstream of TATA[28]. We inferred the approximate position of flexible elements within TFIIB and TFIIF by docking their crystal structures as rigid bodies within our cryo-EM density. Importantly, we find that both the TFIIB linker helix and the TFIIF arm domain align with the promoter melting start site (Supplementary Fig. 13). This arrangement is consistent both with the proposed role of the linker helix of TFIIB in promoter opening[12] and with the crosslinking of the arm domain of RAP74 to the TFIIB linker near the active site[29,30], as well as with the suppression of the TSS defect of TFIIB mutations by a mutant within the arm domain of TFIIF[31]. In our rigid-body fitting, the linker helix of TFIIB overlaps with the DNA in our model, suggesting a rearrangement of the helix relative to the clamp domain at this stage in the PIC assembly. Finally, the tip of the TFIIF arm domain contains seven positively charged residues, whereas four positively charged residues are present on the side of the TFIIB linker helix that faces the DNA (Supplementary Fig. 13). The juxtaposition of these domains within the melting start site is consistent with their direct role in DNA interactions.

The structural features of our Pol II based PIC model are likely conserved with Pol I and Pol III, the two other RNA polymerases in eukaryotes. A side-by-side comparison of our Pol II-based PIC model with a cryo-EM structure of native Pol III agrees with this hypothesis[32] (Supplementary Fig. 14).

Transition to the open-promoter complex

It is well established that the PIC remains stably associated during transcription initiation until Pol II undergoes promoter escape[2]. Preceding this step, however, Pol II needs to transition into an OC in which the melted single-stranded DNA is inserted into the active site. To gain structural insight into the transition from a closed to an open promoter

complex, we generated a “functional mimic” of the PIC in its open conformation by modifying the promoter substrate used to form the closed PIC (Fig. 4a). We replaced the segment of DNA containing the INR element with a 3’-tailed sequence previously used to create an arrested transcription state in yeast Pol II[33]. We matched the arrested position of Pol II on the template exactly to the TSS used in our studies, thereby creating a Pol II-nucleic acid complex containing only ~5 nucleotides at the active site, while still containing upstream core promoter elements available for assembling the rest of the PIC. We found that TFIIE had a higher affinity for the OC, as excess TFIIE had to be used to saturate the closed PIC, but not the open state mimic. Interestingly, excess TFIIE was no longer required in the context of the closed PIC when TFIIH was also included (see later), in agreement with previous studies suggesting cooperative binding of TFIIE and TFIIH within the PIC[2,34,35].

The reconstruction of the OC mimic resembled that of the PIC in the closed conformation, with all the GTFs remaining at identical positions (Fig. 4b and Supplementary Fig. 4e, 5e, 10b–d). This finding is consistent with the prevalent hypothesis that the PIC assembled at the promoter remains intact until promoter escape[2]. In contrast, the downstream DNA adopted a conformation previously observed for elongating Pol II, indicating that the template strand was inserted through the positively charged cleft into the active site[4] (Fig. 4b,c). The single stranded segments are invisible at our resolution or not present (non-template strand). As a reasonable model, the bubble depicted has been derived from a previous model based on FRET studies on the yeast system[36].

When the position and orientation of the downstream DNA is compared between the closed PIC and the OC mimic, it is clear that there is a change in orientation concomitant with the insertion of the downstream DNA into the active site (Fig. 4c), indicating that the DNA rotates on a plane as it translates, while maintaining a point of contact between the DNA and RPB5 that corresponds to one of the two contacts present in the closed state (the one downstream of the INR).

Other than the repositioning of the DNA within the active site, two main differences were observed upon comparison of the OC mimic and the closed PIC structures. First, the clamp domain in the open state moves down to engage the open DNA bubble, adopting the conformation observed in the elongation state[37] (Fig. 4d and Supplementary Fig. 11d). Thus, the clamp domain completes an open to closed transition throughout the process of PIC assembly and promoter opening (Supplementary Fig. 11), a cycle also reported for the bacterial system[38]. Second, an additional protein density now extends from the bottom of the clamp and connects to the dimerization domain of TFIIF (Fig. 4e). Rigid body fitting of crystal structures suggests that this density corresponds to the stabilized rudder of Pol II and the arm domain of TFIIF. We propose that these elements interact with each other as the clamp closes down over the melted DNA. Interestingly, this proposed interaction would prevent re-annealing of the melted DNA. The TFIIB linker helix is near this position and likely participates in the promoter melting process as well. This proposal is consistent with our hypothesis that the flexible TFIIB linker helix and the TFIIF arm domain act together in promoter opening (Supplementary Fig. 13). Thus, our structure and pseudo-atomic model provide a possible explanation for the enigmatic role of TFIIF in promoter opening and TSS selection[22,23].

Positioning of XPB for DNA translocation

To gain insight into the natural promoter opening process carried out by TFIIH, we utilized the same purification strategy used for the previous closed PIC subcomplexes but included the purified, endogenous 10-subunit human TFIIH complex as a last step before elution. Given the scarcity of purified human TFIIH, this study was limited to negative stained samples, which require less material. The 3D reconstruction of the TFIIH-containing PIC showed a substantial additional density extending away from Pol II, consistent with the large molecular weight of TFIIH (0.45 MDa) (Fig. 1e, Fig. 5a and Supplementary Fig. 6e). Surprisingly, only two contacts are observed between TFIIH and the rest of the PIC. One is with the Pol II's stalk domain, at the site of interaction with TFIIE. The other contact likely involves the interaction of TFIIH directly with the downstream DNA. While the DNA is not visible in this negative stain reconstruction, its position can be extrapolated from the cryo-EM structure of the PIC containing TFIIE (Fig. 5a).

The CAK subcomplex (Cdk7/CyclinH/Mat1) of TFIIH, which phosphorylates the C-terminal domain (CTD) of RPB1, is missing from our PIC reconstruction based on comparison with a recent EM study of yeast TFIIH[39]. When we analyzed images of free human TFIIH, an additional density that could accommodate the mass of the CAK subcomplex appeared highly mobile, in agreement with the yeast TFIIH data[39] (Supplementary Fig. 15, 16). Interestingly, when this new density, which fits the crystal structures of Cdk7 and Cyclin H, is placed in the context of the full PIC, it faces towards the CTD of RPB1 (Supplementary Fig. 16).

The reconstruction of the TFIIH-containing PIC allowed us to dock the crystal structure of XPD[40] and a homology model of XPB[6] (Fig. 5a and Supplementary Fig. 6e). XPD is positioned in close proximity to TFIIE and the Pol II stalk, but away from DNA, consistent with a scaffolding role in transcription initiation[7]. On the other hand, XPB docked directly on the downstream DNA path, between the +10 and +20 bp position relative to the TSS (Fig. 5a). This position is consistent with previous crosslinking data using purified TFIIH[41], but inconsistent with a recent crosslinking study using overexpressed XPB in extracts, in which XPB was proposed to be positioned closer to the TFIIE WH domains and the INR element[6]. This result might reflect an alternative position of this protein during the assembly of the PIC, a distinct position of XPB on the DNA when out of the TFIIH complex, or the effect of other factors like TFIID and Mediator on PIC organization. The position of XPB we observed within the TFIIH density, together with the movement of the downstream DNA inferred from comparison of our reconstructions of the closed and open states of Pol II, suggests how XPB could act as a DNA translocase. A translocase model for XPB has previously been proposed[6], but our structure now shows XPB positioned further downstream, leaving enough space around the INR element for it to be melted during this process.

We believe that the position of XPB suggests a DNA insertion process in which, as XPB walks on the DNA away from the rest of PIC, the DNA would be translocated in the opposite direction and pushed into the Pol II cleft while maintaining a point of contact with RPB5 (which starts involving the DNA just downstream of the INR in the closed complex).

This happens concomitantly with a rotation of the DNA, with the RPB5 contact likely serving as a pivot point. As XPB walks on the DNA helix, it would generate supertwist that would be relaxed by unwinding. While this unwinding cannot happen in the DNA that is tightly wrapped and stabilized by the TBP-TFIIA-TFIIB-Rap30-protrusion module, it would be facilitated and/or stabilized in the DNA region between the BREd and the INR, where the arm domain of Rap74 makes contact with the Pol II rudder and regions of TFIIB.

Concluding remarks

The combination of structures described here provides unprecedented mechanistic insight into the stepwise assembly of the human PIC, defining key protein-protein and protein-DNA interactions important for PIC function (Fig. 5b). Our structures reveal the location and role of RAP30 WH domain within an essential upstream nucleoprotein subcomplex. Its critical function in structurally stabilizing the whole PIC is highlighted by our direct visualization of the DNA as it positions along the Pol II cleft upon TFIIF binding. We also show a direct interaction between the arm module of the TFIIF dimerization domain and the rudder domain of Pol II upon formation of the OC, leading to a direct mechanistic model of how this TFIIF element facilitates and/or maintains strand separation concomitant with the closing down of the clamp domain of Pol II. Our structures show how two essential factors, TFIIB and TFIIF come together at critical locations for their activity in the context of a full PIC. Our studies also reveal how TFIIH, because of its large size, can simultaneously interact with TFIIE at the base of the Pol II stalk and position XPB on downstream DNA.

Our studies of the closed PIC and an OC mimic illuminate the structural transitions necessary during the process of promoter melting. The apparent movement of downstream DNA observed when comparing the closed PIC and OC structures, together with the positioning of XPB on the downstream DNA, suggests how XPB could act as a DNA translocase to thread approximately 10bp of downstream double stranded DNA into the cleft. This translocating activity would push against the stably bound upstream DNA around the TATA box to induce negative supercoiling near the TSS. We find that the TFIIB linker helix and the TFIIF arm domain align with each other at the promoter melting start site, likely to facilitate the separation of the two strands. Once promoter DNA melting is further extended and the Pol II clamp closes down, the TFIIB linker helix and the TFIIF arm domain work together with the Pol II rudder to maintain the upstream edge of the DNA bubble.

Finally, the arrangement of components within our PIC structure is compatible with existing structural models that include the large, multi-subunit Mediator and TFIID complexes[10]. Future structural studies with Mediator and/or TFIID will yield further insight regarding the regulation of PIC assembly and function. In summary, this work provides the structural framework needed to integrate biochemical and structural data into a unified mechanistic understanding of transcription initiation.

METHODS SUMMARY

Protein expression and purification

TBP, TFIIA, TFIIB, TFIIE, and TFIIF were recombinantly expressed and purified from *Escherichia coli*. Pol II and TFIIH were immunopurified from HeLa cell nuclear extracts[42]. The design of the DNA construct was based on the SCP[43], with a SalI restriction enzyme site introduced downstream of the INR element. PIC complexes were assembled according to an *in vitro* transcription protocol[42] with minor modifications (see extended Supplementary Methods). The reactions were incubated with magnetic streptavidin T1 beads (Invitrogen) and the desired complexes were eluted by SalI digestion.

EM and image analysis

Data collection and image processing were conducted using the Leginon data collection software[44] and the Appion electron microscopy processing environment[45], respectively. Three-dimensional maps were calculated using libraries from the EMAN2 and SPARX software packages[46,47]. Volume segmentation, automatic rigid body docking, figure and movie generation were performed using UCSF Chimera[48].

METHODS

PIC assembly and purification

TBP, TFIIA, TFIIB, TFIIE, and TFIIF were recombinantly expressed and purified from *Escherichia coli*. Pol II, TFIID, and TFIIH were immunopurified from HeLa cell nuclear extracts following previously established protocols[42,51]. The design of the DNA construct was based on the SCP[43], except that a BREu element was introduced upstream of the TATA box[14] and a SalI restriction enzyme site was included downstream of the INR element for purification purposes (Template1, 5'-ACTGGGAAGTCGACCGGTCCGTAGGCACGTCTGCTCGGCTCGAGTGTTCGATCGCGACTGAGGACGAACGCGCCCCCACCCCCTTTTATAGGCGCCCTTC; Nontemplate1, 5'-GAAGGGCGCCTATAAAAGGGGGTGGGGGCGCGTTCGTCCTCAGTCGCGATCGAACACTCGAGCCGAGCAGACGTGCCTACGGACCGGTCGACTTCCCAGT). The nucleic acid scaffold that was used to generate the PIC in the open conformation was designed by modification of the promoter substrate used to form the closed PIC. An RNA-DNA duplex beyond 7 bp has been proposed to be the trigger for TFIIB release and promoter escape[12,52]. Thus, we replaced the segment of DNA containing the INR element with a 3'-tailed sequence previously used to create an arrested transcription state in yeast Pol II[33]. We matched the arrested position of Pol II on the template exactly to the TSS used in our studies, thereby creating a Pol II-nucleic acid complex containing only about 5 nucleotides at the active site while still containing upstream core promoter elements available for assembling the rest of the PIC (Template2, 5'-ACTGGGAAGTCGACCGGTCCGTAGGCACGTCTGCTCGGCTCGAGTGAGCTAGCTTACCTGGTGTTGCTCTAACCCCCACCCCCTTTTATAGGCGCCCTTC; Nontemplate2, 5'-GAAGGGCGCCTATAAAAGGGGGTGGGGGTT; Nontemplate3, 5'-GAGGTAAGCTAGCTCACTCGAGCCGAGCAGACGTGCCTACGGACCGGTCGACTT

CCCAGT). A biotin tag was engineered at the 5' end of both template strands (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 1 h.

PIC in the closed conformation were assembled according to an in vitro transcription protocol[42] with minor modifications. The assembly buffer contained 12 mM HEPES, pH 7.9, 0.12 mM EDTA, 12% glycerol, 8.25 mM $MgCl_2$, 60 mM KCl, 1 mM DTT, 0.05% NP-40, 2.5 ng/μl dI-dC, 10 μM $ZnCl_2$. The following purified proteins and nucleic acids were sequentially added into the assembly buffer: Pol II, TFIIB, TBP/TFIIA, DNA (template1-nontemplate1), TFIIF, and TFIIE at final concentrations of 185 nM, 3.6 μM, 370 nM, 50 nM, 289 nM, and 370 nM, respectively. The assembly reaction was kept at 37°C for an additional 5 min whenever a new factor was added. The reaction was incubated at 28°C for 15 min using a 1:10 dilution of the magnetic streptavidin T1 beads (Invitrogen) which had been equilibrated with the assembly buffer. Following washing of the beads three times using a washing buffer (10 mM HEPES, 10 mM TRIS, pH 7.9, 5% glycerol, 5 mM $MgCl_2$, 50 mM KCl, 1 mM DTT, 0.05% NP-40, 5 μM $ZnCl_2$), TFIIH at a final concentration of 100 nM was incubated with the beads in assembly buffer at 37°C for an additional 5 min. Following a single additional wash of the beads using washing buffer, the desired complex was eluted by incubating the beads at 28°C for 1 h with digestion buffer containing 10 mM HEPES, pH 7.9, 5% glycerol, 10 mM $MgCl_2$, 50 mM KCl, 1 mM DTT, 0.05% NP-40, 1 unit/μl BSA-free SalI-HF (New England Biolabs). The various PIC intermediates were generated by including just the factors of interest during the assembly process described above. For preparing TBP-TFIIA-TFIIB-DNA-PolII-TFIIF-TFIIE complex, extra TFIIE was added afterwards to the purified PIC at a final concentration of 100 nM.

It was not possible for us to reconstitute an open complex using either a mismatch DNA bubble, probably due to failure of efficiently and specifically positioning Pol II on the bubble, or a nucleic acid scaffold containing an RNA primer, probably because an RNA-DNA duplex of over 7 bp in length has been proposed to be the trigger for TFIIB release and promoter escape[12,52]. PIC in the open conformation was assembled similarly, except for the following changes. An arrested Pol II on the open promoter nucleic acid scaffold was first prepared by incubation at 28°C for 1 h of Pol II and DNA (template2-nontemplate3) at final concentrations of 300 nM and 80 nM, respectively, in the arresting buffer containing 12 mM HEPES, pH 7.9, 0.12 mM EDTA, 12% glycerol, 8.25 mM $MgCl_2$, 60 mM KCl, 1 mM DTT, 0.05% NP-40, 2.5 ng/μl dI-dC, 1:100 dilution of RNasin Ribonuclease inhibitor (Promega), and 2 mM CTP. The following purified proteins and nucleic acid were sequentially added into the arrested Pol II reaction above: nontemplate2, TBP/TFIIA, TFIIB, TFIIF, and TFIIE at final concentrations of 200 nM, 370 nM, 3.6 μM, 289 nM, and 370 nM, respectively. The desired open promoter complex was then purified in the same manner as the closed complexes above.

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 5min, then immediately used for EM sample preparation (either negative stain or cryo-plunging).

Electron microscopy

Negative stain samples of PIC complex and of free TFIIH were prepared using 400 mesh copper grid containing a continuous carbon supporting layer. The grid was plasma cleaned for 10 s immediately prior sample deposition using a Solarus plasma cleaner (Gatan) equipped with 75% argon/25% oxygen. An aliquot (3 μl) of the purified sample (~50 nM) was placed onto the grid and allowed to absorb for 5 min at 100% humidity in a homemade humidity chamber kept under very low illumination conditions. It was subsequently stained by five successive 75 μl drops of 2% (w/v) uranyl formate solution, rocking 10 s on each drop followed by blotting till dryness. Data collection was performed using a Tecnai F20 Twin transmission electron microscope operating at 120 keV at a nominal magnification of ×80,000 (1.37 Å per pixel). The data was collected using the Leginon data collection software[44] on a Gatan 4k×4k camera using low-dose procedures (20 e^{-} $Å^{-2}$ exposures) and a range of defocus values (from −0.5 to −1.2 μm). Between 300 and 600 images were acquired for each of the negative stain data sets.

Preparation of PIC samples for cryo-EM observation was carried out using 400 mesh C-flats 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) equipped with 75% argon/25% oxygen gas immediately prior sample deposition. Aliquot (3 μl) of the purified sample (~100 nM) was placed onto the grid and loaded into a Vitrobot (FEI) 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 Tecnai F20 Twin transmission electron microscope operating at 120 keV using a 626 single-tilt cryotransfer system (Gatan). Data were acquired at a nominal magnification of ×100,000 (1.05 Å per pixel) using low-dose procedures (20 e^{-} $Å^{-2}$ exposures) and a range of defocus values (from −1.2 to −2.4 μm). Between 1500 and 3200 images for each of the cryo data sets were collected using the MSI-T application of the Leginon data collection software[44].

Image processing

Negative stain data pre-processing was performed using the Appion processing environment[45]. Particles were automatically selected from the micrographs using a difference of Gaussians (DoG) particle picker[53]. The contract transfer function (CTF) of each micrograph was estimated using both the ACE2 and CTFFind programs during data collection[54,55], the phases were flipped using CTFFind, and particle stacks were extracted using a box size of 256×256 pixels (except for both the TFIIH containing PIC complex and free TFIIH samples, which use 320×320 pixel boxes) from images whose ACE2 confidence value was greater than 0.8, followed by normalization using the XMIPP program to remove pixels which were above or below 4.5 σ of the mean value[56]. The particle stack was binned by a factor of two and two-dimensional classification was conducted using iterative multivariate statistical analysis and multireference alignment analysis (MSA-MRA) within the IMAGIC software[57]. Class averages containing properly assembled complexes were manually selected and re-extracted to create a new particle stack for reconstruction.

Cryo data processing was performed in a similar manner as the negative stain data. Particle stack was extracted using 384×384 pixel box size from phase-flipped images and binned by a factor of two.

Three-dimensional reconstruction

The cryo-negative staining structure of free Pol II[58], after low-pass filtering to 60 Å, was used as initial model for reconstruction of all the negatively stained PIC samples. For reconstruction of the TFIIH containing PIC, which has a substantial extra mass with respect to Pol II, the negative stain reconstruction of the TBP-TFIIA-TFIIB-DNA-Pol II-TFIIF-TFIIE sample, after low-pass filtering to 60 Å, was instead used as the initial reference (Supplementary Fig. S1). For reconstruction of the free TFIIH, the core TFIIH density segmented from the negative stain TBP-TFIIA-TFIIB-DNA-Pol II-TFIIF-TFIIE-TFIIH refined model was used as the initial reference, after low-pass filtering to 60 Å resolution. Three-dimensional reconstruction was conducted using an iterative multi-reference projection-matching approach containing libraries from the EMAN2 and SPARX software packages[46,47] with two identical copies of the initial model as references. This step allowed us to further eliminate contamination, aggregated, or damaged complexes, which became enriched in one of the reconstructions. Refinement began at an angular step of 25° and progressed down to 4° angular increments. At each step, refinement proceeded to the next angular step only once >95% of the particles had a pixel error of <1 pixel. The particle numbers contributing to the final negative stain reconstructions were 11,880 for TBP-TFIIA-TFIIB-DNA-Pol II, 13,770 for as previous plus TFIIF, 15,656 for as previous plus TFIIE, 64,712 for as previous plus TFIIH, and 13,023 for free TFIIH. The resolution of the reconstruction was estimated using the 0.143 Fourier shell correlation (FSC) criterion to be about 15 Å for PIC complexes, and 20 Å for free TFIIH.

Cryo-EM reconstructions were performed in a similar manner. The negative stain reconstruction of TBP-TFIIA-TFIIB-DNA-Pol II-TFIIF was used as the initial reference for all the cryo reconstructions, except for that of TBP-TFIIA-TFIIB-DNA-Pol II, which used its corresponding negative stain model as the initial reference (Supplementary Figs 2–5). All initial models were low-pass filtered to 60 Å resolution. The particle numbers contributing to the final reconstructions were 122,480 for TBP-TFIIA-TFIIB-DNA-Pol II, 43,785 for as previous plus TFIIF, 51,043 for as previous plus TFIIE, and 53,505 for as previous in the open conformation. To dampen low-resolution amplitudes of the final maps, the Fourier amplitudes were adjusted to match an experimental GroEL SAXS curve using the SPIDER software[59]. The estimated resolution, using the 0.143 FSC criterion, was between 11 and 13 Å for the cryo-EM reconstructions of PIC assembly intermediates (Supplementary Figs 2–5).

Local resolution calculation was performed for all reconstructions using the “blocres” function in the Bsoft package[60,61](Supplementary Figs 2–6). Volume segmentation, automatic rigid-body docking, figure and movie generation were performed using UCSF Chimera[48]. The globally bent DNA model was generated using the 3D-DART online server[62] and the 3DNA software package[63].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Carla Inouye for providing us with recombinant TFIIF and TFIIE, Patricia Grob and Tom Houweling for electron microscopy and computer support, respectively, Tom Goddard for help with Chimera, and members of the Nogales lab for technical advice on image processing. We are thankful to James Kadonaga, James Goodrich, and Michael Cianfrocco for their comments on the manuscript. We thank Priscilla Cooper both for biochemical advice and for her comments on the manuscript. This work was funded by NIGMS (GM63072 to E.N) and by NCI (CA127364 to D.T). E. N. is a Howard Hughes Medical Institute Investigator.

References

1. Matsui T, Segall J, Weil PA, Roeder RG. Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. J Biol Chem. 1980; 255:11992–11996. [PubMed: 7440580]
2. Roeder RG. The role of general initiation factors in transcription by RNA polymerase II. Trends Biochem Sci. 1996; 21:327–335. [PubMed: 8870495]
3. Goodrich JA, Cutler G, Tjian R. Contacts in context: promoter specificity and macromolecular interactions in transcription. Cell. 1996; 84:825–830. [PubMed: 8601306]
4. Kornberg RD. The molecular basis of eukaryotic transcription. Proc Natl Acad Sci U S A. 2007; 104:12955–12961. [PubMed: 17670940]
5. Cramer P, et al. Structure of eukaryotic RNA polymerases. Annu Rev Biophys. 2008; 37:337–352. [PubMed: 18573085]
6. Grunberg 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]
7. Thomas MC, Chiang CM. The general transcription machinery and general cofactors. Crit Rev Biochem Mol Biol. 2006; 41:105–178. [PubMed: 16858867]
8. Andel F 3rd, Ladurner AG, Inouye C, Tjian R, Nogales E. Three-dimensional structure of the human TFIID-IIA-IIB complex. Science. 1999; 286:2153–2156. [PubMed: 10591646]
9. Chung WH, et al. RNA polymerase II/TFIIF structure and conserved organization of the initiation complex. Mol Cell. 2003; 12:1003–1013. [PubMed: 14580350]
10. Bernecky C, Grob P, Ebmeier CC, Nogales E, Taatjes DJ. Molecular architecture of the human Mediator-RNA polymerase II-TFIIF assembly. PLoS Biol. 2011; 9:e1000603. [PubMed: 21468301]
11. 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]
12. Kostrewa D, et al. RNA polymerase II-TFIIB structure and mechanism of transcription initiation. Nature. 2009; 462:323–330. [PubMed: 19820686]
13. 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]
14. Tsai FT, Sigler PB. Structural basis of preinitiation complex assembly on human pol II promoters. EMBO J. 2000; 19:25–36. [PubMed: 10619841]
15. Sainsbury S, Niesser J, Cramer P. Structure and function of the initially transcribing RNA polymerase II-TFIIB complex. Nature. 2012
16. 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]
17. 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]

18. 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]
19. Robert F, Forget D, Li J, Greenblatt J, Coulombe B. Localization of subunits of transcription factors IIE and IIF immediately upstream of the transcriptional initiation site of the adenovirus major late promoter. J Biol Chem. 1996; 271:8517–8520. [PubMed: 8621472]
20. Tyree CM, et al. Identification of a minimal set of proteins that is sufficient for accurate initiation of transcription by RNA polymerase II. Genes Dev. 1993; 7:1254–1265. [PubMed: 8319911]
21. Tan S, Garrett KP, Conaway RC, Conaway JW. Cryptic DNA-binding domain in the C terminus of RNA polymerase II general transcription factor RAP30. Proc Natl Acad Sci U S A. 1994; 91:9808–9812. [PubMed: 7937895]
22. Ghazy MA, Brodie SA, Ammerman ML, Ziegler LM, Ponticelli AS. Amino acid substitutions in yeast TFIIF confer upstream shifts in transcription initiation and altered interaction with RNA polymerase II. Mol Cell Biol. 2004; 24:10975–10985. [PubMed: 15572698]
23. Yan Q, Moreland RJ, Conaway JW, Conaway RC. Dual roles for transcription factor IIF in promoter escape by RNA polymerase II. J Biol Chem. 1999; 274:35668–35675. [PubMed: 10585446]
24. Forget D, et al. RAP74 induces promoter contacts by RNA polymerase II upstream and downstream of a DNA bend centered on the TATA box. Proc Natl Acad Sci U S A. 1997; 94:7150–7155. [PubMed: 9207059]
25. Orlicky SM, Tran PT, Sayre MH, Edwards AM. Dissociable Rpb4-Rpb7 subassembly of rna polymerase II binds to single-strand nucleic acid and mediates a post-recruitment step in transcription initiation. J Biol Chem. 2001; 276:10097–10102. [PubMed: 11087726]
26. Grohmann D, et al. The initiation factor TFE and the elongation factor Spt4/5 compete for the RNAP clamp during transcription initiation and elongation. Mol Cell. 2011; 43:263–274. [PubMed: 21777815]
27. 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 U S A. 1991; 88:7509–7513. [PubMed: 1881889]
28. Giardina C, Lis JT. DNA melting on yeast RNA polymerase II promoters. Science. 1993; 261:759–762. [PubMed: 8342041]
29. Chen HT, Hahn S. Mapping the location of TFIIB within the RNA polymerase II transcription preinitiation complex: a model for the structure of the PIC. Cell. 2004; 119:169–180. [PubMed: 15479635]
30. Freire-Picos MA, Krishnamurthy S, Sun ZW, Hampsey M. Evidence that the Tfg1/Tfg2 dimer interface of TFIIF lies near the active center of the RNA polymerase II initiation complex. Nucleic Acids Res. 2005; 33:5045–5052. [PubMed: 16147988]
31. Sun ZW, Hampsey M. Identification of the gene (SSU71/TFG1) encoding the largest subunit of transcription factor TFIIF as a suppressor of a TFIIB mutation in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A. 1995; 92:3127–3131. [PubMed: 7724527]
32. Fernandez-Tornero C, et al. Conformational flexibility of RNA polymerase III during transcriptional elongation. EMBO J. 2010; 29:3762–3772. [PubMed: 20967027]
33. Cheung AC, Cramer P. Structural basis of RNA polymerase II backtracking, arrest and reactivation. Nature. 2011; 471:249–253. [PubMed: 21346759]
34. Goodrich JA, Tjian R. Transcription factors IIE and IIH and ATP hydrolysis direct promoter clearance by RNA polymerase II. Cell. 1994; 77:145–156. [PubMed: 8156590]
35. Conaway RC, Conaway JW. General initiation factors for RNA polymerase II. Annu Rev Biochem. 1993; 62:161–190. [PubMed: 8352588]
36. Andrecka J, et al. Nano positioning system reveals the course of upstream and nontemplate DNA within the RNA polymerase II elongation complex. Nucleic Acids Res. 2009; 37:5803–5809. [PubMed: 19620213]
37. Gnatt AL, Cramer P, Fu J, Bushnell DA, Kornberg RD. Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 A resolution. Science. 2001; 292:1876–1882. [PubMed: 11313499]

38. Chakraborty A, et al. Opening and closing of the bacterial RNA polymerase clamp. Science. 2012; 337:591–595. [PubMed: 22859489]
39. Gibbons BJ, et al. Subunit architecture of general transcription factor TFIIH. Proc Natl Acad Sci U S A. 2012; 109:1949–1954. [PubMed: 22308316]
40. 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]
41. Kim TK, Ebright RH, Reinberg D. Mechanism of ATP-dependent promoter melting by transcription factor IIH. Science. 2000; 288:1418–1422. [PubMed: 10827951]
42. Revyakin A, et al. Transcription initiation by human RNA polymerase II visualized at single-molecule resolution. Genes Dev. 2012; 26:1691–1702. [PubMed: 22810624]
43. 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]
44. Suloway C, et al. Automated molecular microscopy: the new Leginon system. J Struct Biol. 2005; 151:41–60. [PubMed: 15890530]
45. Lander GC, et al. Appion: an integrated, database-driven pipeline to facilitate EM image processing. J Struct Biol. 2009; 166:95–102. [PubMed: 19263523]
46. Tang G, et al. EMAN2: an extensible image processing suite for electron microscopy. J Struct Biol. 2007; 157:38–46. [PubMed: 16859925]
47. Hohn M, et al. SPARX, a new environment for Cryo-EM image processing. J Struct Biol. 2007; 157:47–55. [PubMed: 16931051]
48. Goddard TD, Huang CC, Ferrin TE. Visualizing density maps with UCSF Chimera. J Struct Biol. 2007; 157:281–287. [PubMed: 16963278]
49. 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 U S A. 1998; 95:9117–9122. [PubMed: 9689043]
50. Chen HT, Warfield L, Hahn S. The positions of TFIIF and TFIIE in the RNA polymerase II transcription preinitiation complex. Nat Struct Mol Biol. 2007; 14:696–703. [PubMed: 17632521]
51. Knuesel MT, Meyer KD, Bernecky C, Taatjes DJ. The human CDK8 subcomplex is a molecular switch that controls Mediator coactivator function. Genes Dev. 2009; 23:439–451. [PubMed: 19240132]
52. 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]
53. 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]
54. Mallick SP, Carragher B, Potter CS, Kriegman DJ. ACE: automated CTF estimation. Ultramicroscopy. 2005; 104:8–29. [PubMed: 15935913]
55. Mindell JA, Grigorieff N. Accurate determination of local defocus and specimen tilt in electron microscopy. J Struct Biol. 2003; 142:334–347. [PubMed: 12781660]
56. Sorzano CO, et al. XMIPP: a new generation of an open-source image processing package for electron microscopy. J Struct Biol. 2004; 148:194–204. [PubMed: 15477099]
57. van Heel M, Harauz G, Orlova EV, Schmidt R, Schatz M. A new generation of the IMAGIC image processing system. J Struct Biol. 1996; 116:17–24. [PubMed: 8742718]
58. Kostek SA, et al. Molecular architecture and conformational flexibility of human RNA polymerase II. Structure. 2006; 14:1691–1700. [PubMed: 17098194]
59. Frank J, et al. SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. J Struct Biol. 1996; 116:190–199. [PubMed: 8742743]
60. Heymann JB, Belnap DM. Bsoft: image processing and molecular modeling for electron microscopy. J Struct Biol. 2007; 157:3–18. [PubMed: 17011211]
61. Lander GC, et al. Complete subunit architecture of the proteasome regulatory particle. Nature. 2012; 482:186–191. [PubMed: 22237024]
62. van Dijk M, Bonvin AM. 3D-DART: a DNA structure modelling server. Nucleic Acids Res. 2009; 37:W235–W239. [PubMed: 19417072]

63. Lu XJ, Olson WK. 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. Nat Protoc. 2008; 3:1213–1227. [PubMed: 18600227]

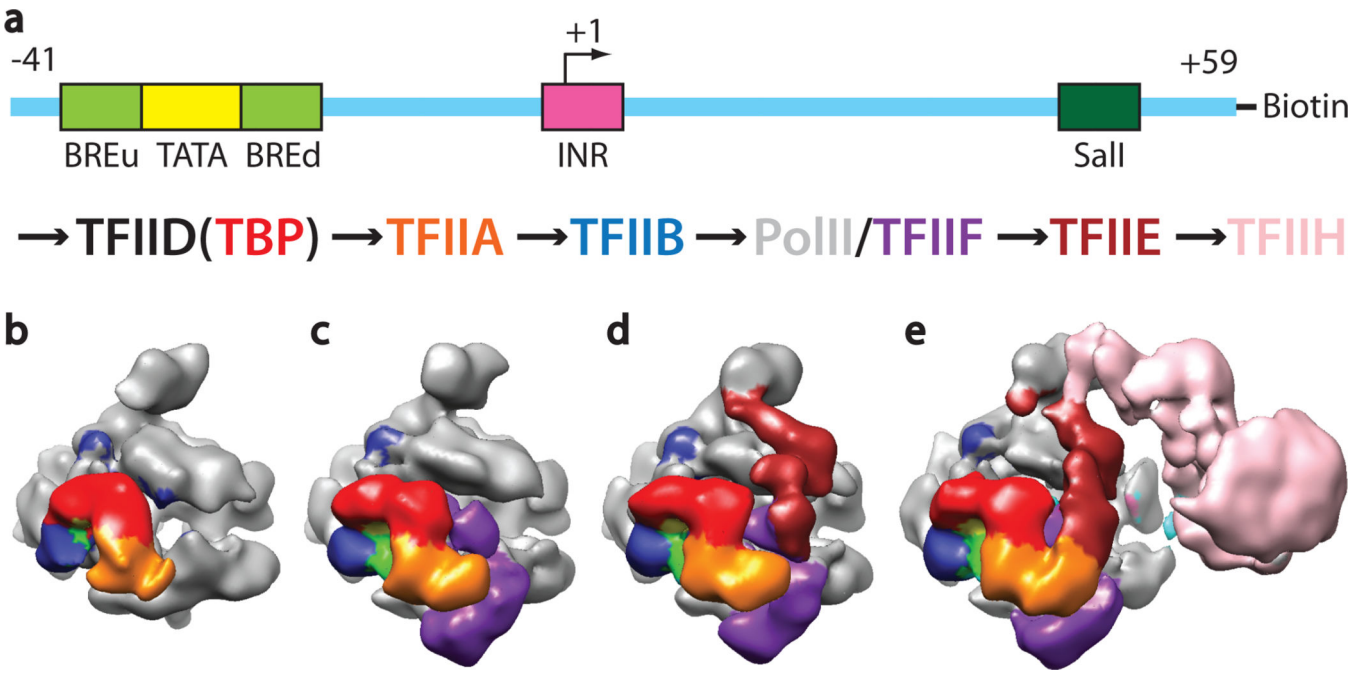

Figure 1. Stepwise assembly of the human PIC

(**a**) Reconstitution strategy for human PIC by sequential assembly. Schematic of the DNA highlighting the relative positions of the core promoter elements used and SalI restriction site (top). Color scheme for the components of the PIC (bottom). Negative stain reconstructions of PIC assembly intermediates for TBP-TFIIA-TFIIB-DNA-Pol II (**b**), plus TFIIF (**c**), plus TFIIE (**d**), and plus TFIIH (**e**).

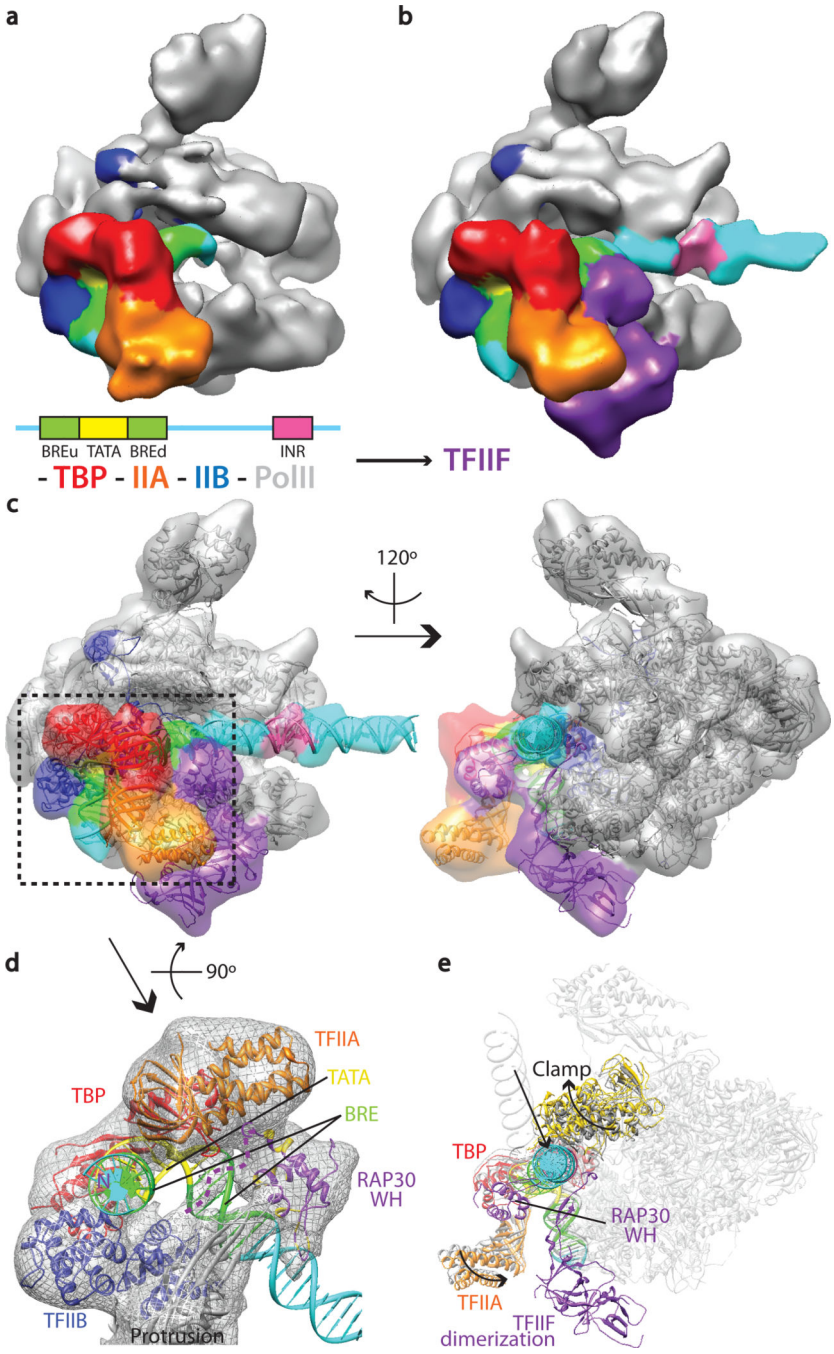

Figure 2. TFIIF engagement triggers a concerted conformational change in the PIC

The positions of TBP, TFIIA, TFIIB and DNA promoter elements relative to Pol II are indicated in the cryo-EM reconstruction in the absence (**a**) or presence of TFIIF (**b**). TFIIF stabilizes promoter DNA (cyan). (**c**) Crystal structures for TBP-TFIIA-DNA (PDB ID: 1NVP), TBP-TFIIB-DNA (PDB ID: 1C9B), Pol II-TFIIB (PDB ID: 4BBR), RAP30/74 dimerization domain (PDB ID: 1F3U[16]), RAP30 WH domain (PDB ID: 1BBY[49]), and modeled B-form DNA (globally bent between −23 and +7 by 18°) are shown docked into transparent EM densities. The mobile clamp of Pol II is docked as a separate domain of Pol

II. (**d**) Bottom view showing the presence of a nucleoprotein complex by the upstream core promoter elements. The DNA densities have been segmented out for clarity. Perturbed residues within the RAP30 WH domain during DNA titrations are colored in gold[49]. The possible path for RAP30 N-terminus is highlighted with dotted purple lines. DNA is shown in ribbon representation. (**e**) Position of moving structural elements before (grey) and after (colored) TFIIF binding. The rest of the PIC components are shown in transparency and major structural rearrangements are depicted by arrows.

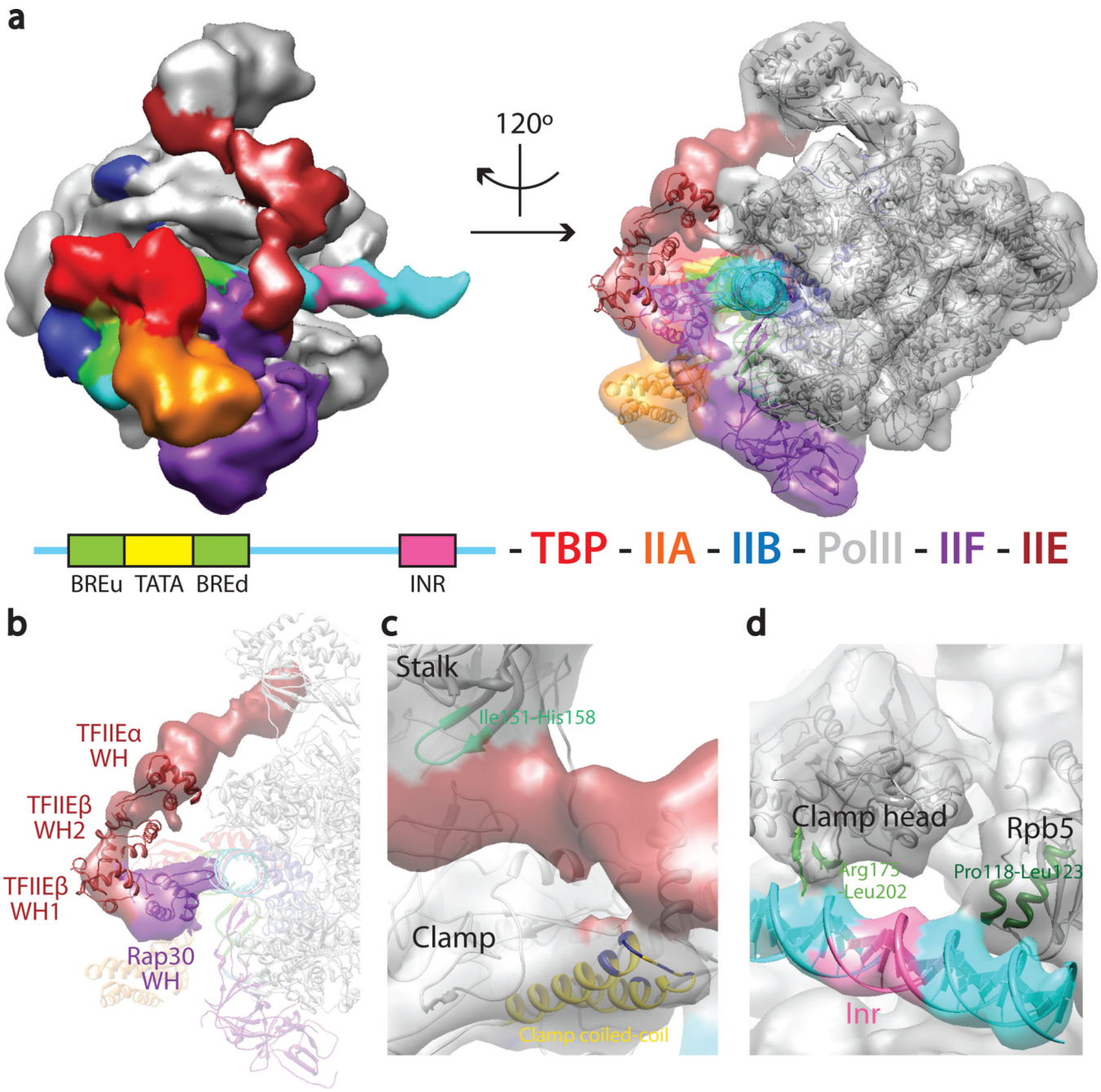

Figure 3. Stabilization of the PIC in the closed conformation by TFIIE

(**a**) Segmentation of the cryo-EM reconstruction of human PIC containing TFIIE (left), and docking of existing crystal structures (right). (**b**) Same view as the right panel in (a) depicting the cryo-EM density corresponding to TFIIE and the RAP30 WH domain. A chain of four WH domains formed by the C-terminus of RAP30 and both subunits of TFIIE (modeled based on cross-linking data[6]) can be roughly fitted into the cryo-EM density. (**c**) Regions of contact between TFIIE and Pol II. The clamp and stalk domain of Pol II are shown in grey, TFIIE in maroon. The rest of the PIC components are shown in transparency. The clamp coiled-coil domain is shown in gold and the residues that crosslinks to TFIIE[50]

are colored in navy blue. (**d**) EM density corresponding to promoter DNA is shown together with the PIC ribbon model. Structural elements making direct contacts with the promoter DNA are depicted.

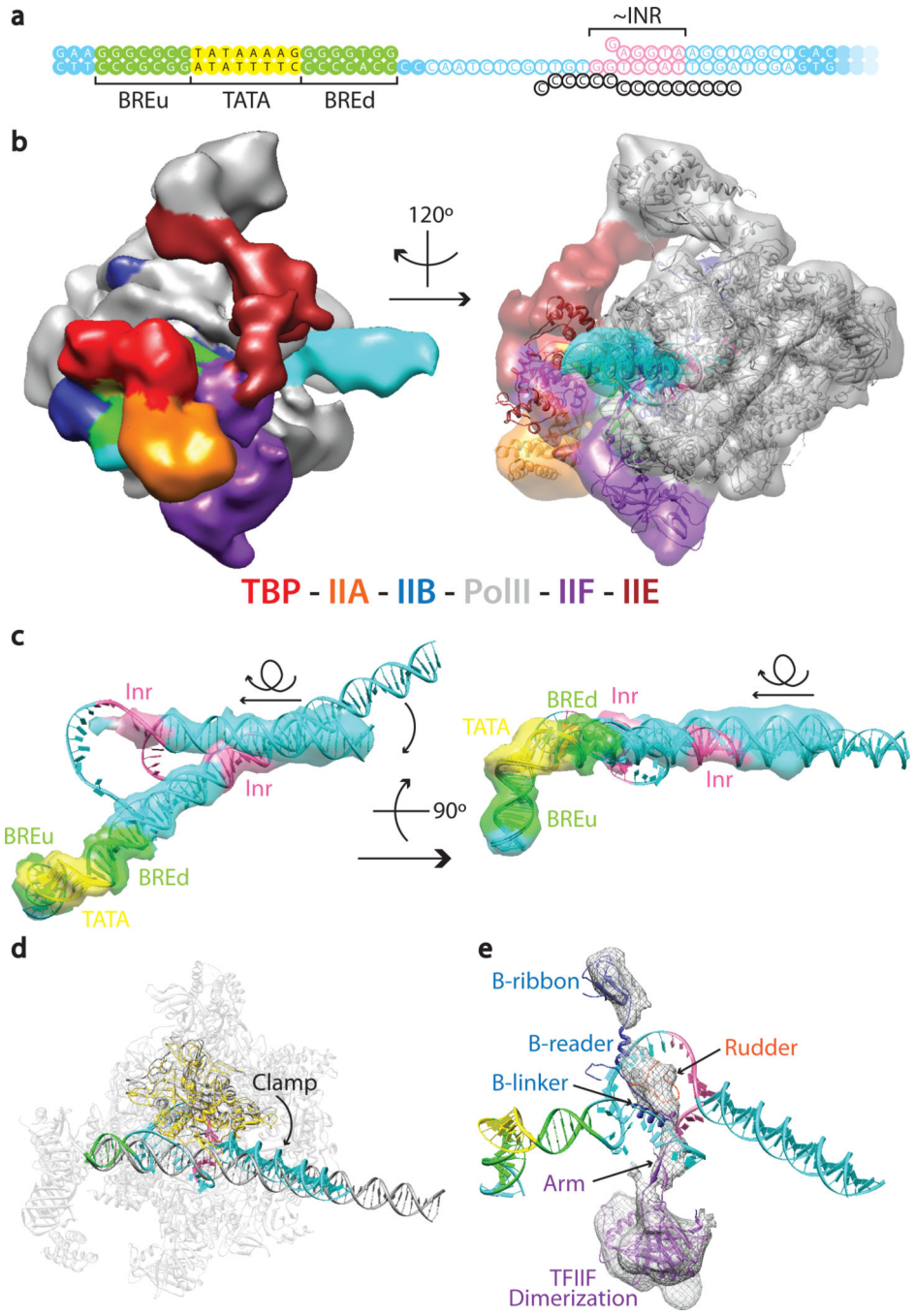

Figure 4. Conformational rearrangements of the PIC upon promoter opening

a) Nucleic acid scaffold used to generate a mimic of the OC. Filled and open circles correspond, respectively, to the core promoter used in Fig. 1a, and to a replacement sequence containing a 3'-tailed sequence previously designed to generate an arrested Pol II[33]. The schematic indicates in pink the position that would correspond to the INR. (**b**) Segmentation of the cryo-EM reconstruction of the human PIC in the open state (left), and docking of existing crystal structures, together with a modeled DNA bubble[12] (right). GTFs adopt the same architecture as in the closed PIC. (**c**) EM densities of promoter DNA in the

closed PIC and OC were segmented and superimposed with respect to Pol II. Movements of the DNA between the two states are shown by arrows. The rotation accompanying translocation of the downstream DNA occurs within the plane of the view shown on the left panel (and thus perpendicular to view shown on the left). (**d**) Structures, before (grey) and after (colored) promoter opening, are shown using fitted crystal structures. The Pol II clamp comes down over the open bubble, in a conformation similar to that seen for the elongation state[37]. (**e**) Segmented cryo-EM density showing the now visible arm domain of TFIIF connecting the rest of TFIIF with the rudder of RPB1. Crystal structures of TFIIB and TFIIF were docked as rigid bodies into the cryo-EM density.

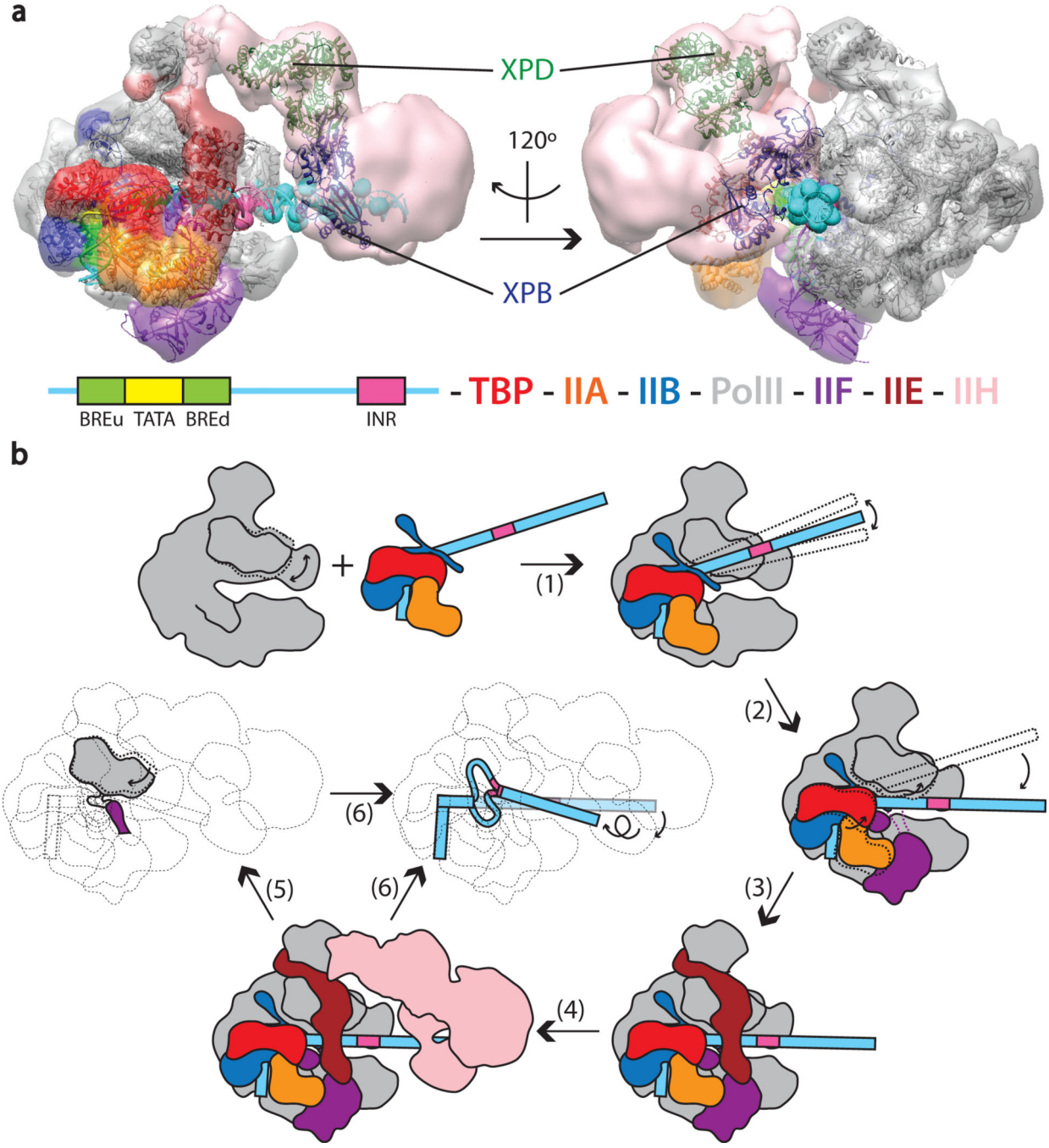

Figure 5. Positioning of TFIIH helicases and model of PIC assembly and promoter opening

(**a**) Negative-stain structure of the full human PIC in the closed conformation (as in Fig. 1e). EM density corresponding to the TFIIH core complex, lacking any visible CAK sub-complex, is colored in pink. A homology model for XPB[6] (navy blue) and the crystal structure of XPD (PDB ID: 3CRV, dark green) are shown docked into the core TFIIH density. The docking suggests different roles for the XPB and XPD helicases in promoter opening. DNA phosphates crosslinked to XPB are indicated by pink or cyan spheres[41]. (**b**) Schematic of PIC assembly and promoter melting. Pol II is recruited through interaction

with TFIIB to the promoter, which is engaged by TBP–TFIIA–TFIIB (**1**). TFIIF stabilizes the TBP-TFIIA-IIB-Pol II protrusion interaction hub and also positions the downstream DNA onto the cleft, forcing the clamp to swing into a slightly open state (**2**). TFIIE binding further stabilizes the PIC by interacting with the Pol II stalk, the clamp, and with TFIIF on the other side of the Pol II cleft (**3**). The TFIIE-containing PIC serves as the platform for TFIIH binding and correctly positions XPB downstream of the INR element (**4**). During strand separation, the clamp domain starts to swing down. The arm domain of TFIIF comes close to the Pol II rudder and the TFIIB B-linker. Stabilization of these interactions forms a physical barrier for DNA re-annealing (**5**). During promoter opening, the translocase activity of XPB would “screw in” the DNA toward the Pol II active site, leading to a Pol II open-promoter state ready for RNA synthesis.