Eukaryotic transcription initiation requires the assembly of general transcription factors into a pre-initiation complex that ensures the accurate loading of RNA polymerase II (Pol II) at the transcription start site. The molecular mechanism and function of this assembly have remained elusive due to lack of structural information. Here 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 pre-initiation complex and the open-promoter complex, and to regulate start-site selection. Comparison of open versus closed pre-initiation complexes, combined with the localization of the TFIIH helicases XPD and XPB, support a DNA translocation model of XPB and explain 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-megadalton (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 escape1–3. Despite recent structural advances on Pol II4,5 and subcomplexes of the PIC6, 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 promoters2–7. TFIIA and TFIIB are then recruited, further stabilizing the interaction between TBP and promoter DNA. Next, Pol II, probably 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 PIC2–5.

Structural characterization of PIC assemblies is challenging and has been limited to a small number of electron microscopy (EM) studies4–10. 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 conformation6,9,11,12. In spite of this progress, important questions remain unanswered, such as how TFIIIB and TFIIH serve complementary roles during the promoter opening process or how TFIIH positions TFIIH in a configuration capable of melting the DNA.

Here we present cryo-EM snapshots of PIC intermediates during sequential assembly (see Supplementary Video 1). 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. Together, our structures provide unprecedented insights into the molecular assembly, organizational and functional roles of different GTFs during transcription initiation.

**Stepwise assembly and visualization of human PIC**

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, BREu and BREd (upstream and downstream TFIIB recognition elements, respectively), and INR (initiator element) 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 charged 2-MDa 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 charged complexes.

**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). Colour scheme for the components of the PIC is shown at the bottom. b–e, Negative stain reconstructions of PIC assembly intermediates for TBP–TFIIA–TFIID–DNA–Pol II (b), plus TFIIF (c), plus TFIIIE (d), and plus TFIIH (e).
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 images 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\(^\text{13}\), TBP–TFIIA–DNA\(^\text{15}\) and yeast Pol II–TFIIA\(^\text{11,12,15}\) 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 proposed previously, based on superimposing crystal structures using the common protein as an anchor point\(^\text{12}\). Our data show that a simple pivoting of the carboxy-terminal cyclin fold domain of TFIIA around the amino-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 TFIIA and Pol II (Supplementary Fig. 7). This small discrepancy with the previous piece-wise model is unlikely to be 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 TFIIIF on PIC assembly

According to the sequential assembly pathway, TFIIIF is recruited to the promoter in association with Pol II\(^\text{2}\). To understand its structural role during PIC assembly, we added TFIIIF separately to our reconstituted system. By comparing the cryo-EM structures of PIC subcomplexes in the absence and presence of TFIIIF, 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 TFIIIF 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\(^\text{12}\) (Supplementary Fig. 8). Thus, TFIIIF is required for the engagement of DNA by Pol II within the context of a closed PIC.

The crystal structure of the human TFIIIF dimerization domain\(^\text{16}\) could be unambiguously fitted into the new density ascribed to TFIIIF by the lobe domain of Pol II using rigid-body docking, only slightly shifted from previous models that were based on crosslinking data\(^\text{17,18}\) (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\(^\text{16}\), indicating that it is mobile at this stage of PIC assembly. A small clash between the RPB2 lobe and the RAP30 WH domain is consistent with the RAP30 WH domain; RAP30 has been shown to crosslink to BREd\(^\text{19}\) and its C-terminal WH domain has been identified to be in direct contact with the protrusion domain of Pol II\(^\text{18}\); (3) RAP30, rather than RAP74, is required for the engagement of DNA by Pol II within the context of a closed PIC. The addition of TFIIIF 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\(^\text{12}\) (Supplementary Fig. 8). Thus, TFIIIF is required for the engagement of DNA by Pol II within the context of a closed PIC.

The crystal structure of the human TFIIIF dimerization domain\(^\text{16}\) could be unambiguously fitted into the new density ascribed to TFIIIF by the lobe domain of Pol II using rigid-body docking, only slightly shifted from previous models that were based on crosslinking data\(^\text{17,18}\) (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\(^\text{16}\), indicating that it is mobile at this stage of PIC assembly. A small clash between the RPB2 lobe and the RAP30 WH domain is consistent with the RAP30 WH domain; RAP30 has been shown to crosslink to BREd\(^\text{19}\) and its C-terminal WH domain has been identified to be in direct contact with the protrusion domain of Pol II\(^\text{18}\); (3) RAP30, rather than RAP74, is required for accurate transcription initiation\(^\text{20}\) and deletion mutants of RAP30’s WH domain are lethal in yeast\(^\text{18,21}\). This WH domain therefore contributes to a unique nucleoprotein complex formed by TBP, TFIIA and the TFIIF 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 and Supplementary Fig. 10).

Comparison of structures shown in Fig. 2a and b strongly suggests that the overall effect of TFIIIF on the assembling PIC is a clear stabilization of the DNA along the Pol II cleft. Given its location, we propose that the RAP30 WH domain has 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

---

**Figure 2** | TFIIIF engagement triggers a concerted conformational change in the PIC. **a, b,** The positions of TBP, TFIIA, TFIIB, DNA promoter elements relative to Pol II are indicated in the cryo-EM reconstruction in the absence (a) or presence of TFIIIF (b). TFIIIF stabilizes promoter DNA (cyan). **c,** Crystal structures for TBP–TFIIA–DNA (PDB accession INVP), TBP–TFIIA–DNA (PDB 1C9B), Pol II–TFIIA (PDB 4BBR), RAP30/74 dimerization domain (PDB 1F3U\(^\text{16}\)), RAP30 WH domain (PDB 1BBY\(^\text{19}\)) and formed 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. **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 coloured in gold\(^\text{16}\). The possible path for the 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 (coloured) TFIIIF binding. The rest of the PIC components are shown in transparency and major structural rearrangements are depicted by arrows required for accurate transcription initiation\(^\text{20}\) and deletion mutants of RAP30’s WH domain are lethal in yeast\(^\text{18,21}\). This WH domain therefore contributes to a unique nucleoprotein complex formed by TBP, TFIIA and the TFIIF 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 and Supplementary Fig. 10).

Comparison of structures shown in Fig. 2a and b strongly suggests that the overall effect of TFIIIF on the assembling PIC is a clear stabilization of the DNA along the Pol II cleft. Given its location, we propose that the RAP30 WH domain has 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
the structural features of our Pol II–based PIC model are probably 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 (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. 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. We matched the 11 Å resolution reconstruction of the PIC containing TFIIE to the growing PIC resulted in new protein density that connects TFIIH with the stalk domain of Pol II (Fig. 3a). The resolution of this reconstruction (11 Å) was the highest obtained for any of the complexes studied, indicating that TFIIH stabilizes the PIC. The density corresponding to TFIIH, 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 TFIIH structure (Fig. 3a, b). One end of TFIIH 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 the deletion of which completely abolished transcription. Also consistent with the positioning of TFIIH in our PIC structure, a zinc ribbon domain within the archaeal homologue of TFIIH was found to be located near the base of the stalk domain of the polymerase. Away from the stalk, the TFIIH density contacts the Pol II clamp domain to interact ultimately with the WH domain of TFIIH. A model of the three WH domains within TFIIH interacting with elements of the clamp head has been proposed based on crosslinking studies (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 TFIIH 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 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 whereas 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 after TFIIH addition is probably due to the interaction of the DNA with the clamp head β-sheet (Fig. 3d and Supplementary Fig. 11b, c).

The spacing between the TATA box and the TSS can vary between species, whereas the region within promoter DNA that is melted during transcription initiation is ~20 base pairs (bp) downstream of TATA. We inferred the approximate position of flexible elements within TFIIA and TFIIH by dock- ing their crystal structures as rigid bodies within our cryo–EM density. Importantly, we find that both the TFIIA linker helix and the TFIIH 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 TFIIA in promoter opening and with the crosslinking of the arm domain of RAP74 to the TFIIA linker near the active site, as well as with the suppression of the TSS defect of TFIIA mutations by a mutant within the arm domain of TFIIH. In our rigid-body fitting, the linker helix of TFIIH 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 TFIIH arm domain contains seven positively charged residues, whereas four positively charged residues are present on the side of the TFIIA 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 probably 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 (Supplementary Fig. 14).
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 TFIHH was also included (see below), in agreement with previous studies suggesting cooperative binding of TFIIE and TFIHH within the PIC$^{34,35}$.

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 Figs 4e, 5e and 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$^9$ (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 in Fig. 4 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$^9$ (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$^{37}$. 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 reannealing of the melted DNA. The TFIIF linker helix is near this position and probably 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$^{2,23}$.

**Positioning of XPB for DNA translocation**

To gain insight into the natural promoter opening process carried out by TFIHH, we used the same purification strategy used for the previous closed PIC subcomplexes but included the purified, endogenous 10-subunit human TFIHH complex as a last step before elution. Given the scarcity of purified human TFIHH, this study was limited to negative stained samples, which require less material. The three-dimensional reconstruction of the TFIHH-containing PIC showed a substantial additional density extending away from Pol II, consistent with the large molecular mass of TFIHH (0.45 MDa) (Figs 1e and 5a and Supplementary Fig. 6e). Surprisingly, only two contacts are observed between TFIHH 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 probably involves the interaction of TFIHH directly with the downstream DNA. Although 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–cyclin H–MAT1) of TFIHH, 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 TFIHH$^{38}$. When we analysed images of free human TFIHH, an additional density that could accommodate the mass of the CAK subcomplex appeared highly mobile, in agreement with the yeast TFIHH data$^{39}$ (Supplementary Figs 15 and 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).

---

**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$^9$. 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 modelled DNA bubble$^{37}$ (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 (coloured) 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$^9$. 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.
The reconstitution of the TFIIH-containing PIC allowed us to dock the crystal structure of XPD\(^{30}\) and a homology model of XPB\(^{3}\) (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 XBP in extracts, in which XPB was proposed to be positioned closer to the TFIIH\(^{3}\) WH domains and the INR element.\(^7\) 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 that 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,\(^8\) 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 probably serving as a pivot point. As XPB walks on the DNA helix, it would generate supertwist that would be relaxed by unwinding. Although this unwinding cannot happen in the DNA that is tightly wrapped and stabilized by the TBP–TFIIE–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 and Supplementary Video 1). 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 is positioned along the Pol II cleft upon TFIIH binding. We also show a direct interaction between the arm module of the TFIIH dimerization domain and the rudder domain of Pol II upon formation of the OC, leading to a direct mechanistic model of how this TFIIH 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, TFIIH 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 10 bp 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 TFIID linker helix and the TFIIH arm domain align with each other at the promoter melting start site, probably to facilitate the separation of the two strands. Once promoter DNA melting is further extended and the Pol II clamp closes down, the TFIIH linker helix and the TFIIH 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 TFIIH complexes.\(^9\) Future structural studies with mediator and/or TFIIH 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, TFIIH, TFIE and TFIIF were recombinantly expressed and purified from *Escherichia coli*. Pol II and TFIIH were immunopurified from HeLa cell nuclear extracts.\(^{8}\) The design of the DNA construct was based on the SCP\(^{44}\), with a Sall restriction enzyme site introduced.
downstream of the INR element. PIC complexes were assembled according to an in vitro transcription protocol \( \text{with minor modifications (see Methods).} \) The reactions were incubated with magnetic streptavidin T1 beads (Invitrogen) and the desired complexes were eluted by S\( \text{al digestion. EM and image analysis. Data collection and image} \) processing were conducted using the Phenix data collection software \( \text{and the Appion electron microscopy processing} \) environment \( \text{respectively. Three-dimensional maps were calculated} \) using libraries from the EMAN2 and SPARX software packages \( \text{Volume segmentation, automated rigid-body docking, figure and movie} \) generation were performed using UCSF Chimera. \text{Full Methods and any associated} \) references are available in the online version of the paper.

Received 26 September 2012; accepted 7 February 2013.

Published online 27 February 2013.

1. Matsui, T., Segal, J., Wei, P. A. & Roeder, R. G. Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. J. Biol. Chem. 255, 11992–11996 (1980).
2. Roeder, R. G. The role of general initiation factors in transcription by RNA polymerase II. Trends Biochem. Sci. 21, 327–335 (1996).
3. Goodrich, J. A., Cutler, G. & Tjian, R. Contacts in context: promoter specificity and nonspecific macromolecular interactions in transcription. Cell 94, 825–830 (1996).
4. Kornberg, R. D. The molecular basis of eukaryotic transcription. Proc. Natl Acad. Sci. USA 104, 12959–12961 (2007).
5. Cramer, P. et al. Amino acid substitutions in yeast TFIIF confer upstream shifts in transcription initiation and altered interaction with RNA polymerase II. Mol. Cell 9, 10975–10985 (2004).
6. Chakraborty, A. et al. Transcription factors IIE and IIH and ATP hydrolysis direct promoter closure by RNA polymerase II. Cell 77, 145–156 (1994).
7. Revyakin, A. et al. Automated molecular microscopy: the new Leginon system. J. Struct. Biol. 119, 71–90 (2001).
8. Goddard, T. D., Huang, C. C. & Ferrin, T. E. Visualizing density maps with UCSF Chimera. Curr. Protoc. Nucleic Acid Chem. Sect. 27, 1–27 (2003).
9. Grohmann, D. et al. Identification of a minimal set of proteins that is sufficient for transcription preinitiation complex: a model for the structure of the PIC. Cell 119, 169–180 (2004).
10. Sainsbury, S., Niesser, J. & Cramer, P. Structure and function of the initially cross-linked TBP–TFIIA–TFIIB–DNA–Pol II–TFIIF–TFIIE in the OC mimic state). Negative stain EM density maps have been assigned to the 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). Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to E.N. (ENogales@lbl.gov).
METHODS
PIC assembly and purification. TP, TFIH, TFIH, TFIE and TFIIF were recombined, purified and infected from E. coli. The renatured TFIIF was immuno-purified from HeLa cell nuclear extracts following previously established protocols. The design of the DNA construct was based on the SCP4, except that a 3′-tailed sequence previously used to create an arrested transcription state in yeast Pol II. We matched the arrested position of Pol II on the template exactly to the TSS used in our studies, thereby creating a Pol II–nuclear acid complex containing only Pol II 5′ nucleotides at the active site while still containing upstream core promoter elements available for assembling the rest of the PIC (template2, 5′-ACTGGGAATGCTGCGCTTCGTGATGTCGACCGGATGAGACGTGCCTACGGACCGGTCGACT-3′; non-template1, 5′-GAAGGGCGCCTATAAAAGGGGGTGCCCGGGTTCTGCTCTCAGTCCGATGTCGACCGGATGAGACGTGCCTACGGACCGGTCGACT-3′; non-template3, 5′-GAAGGGCGCCTATAAAAGGGGGTGCCCGGGTTCTGCTCTCAGTCCGATGTCGACCGGATGAGACGTGCCTACGGACCGGTCGACT-3′). The nucleic acid scaffold that was used to generate the PIC in the open conformation was developed by modification of the promoter substate used to form the closed PIC. An RNA–DNA duplex beyond 7 bp has been proposed to be the trigger for TFIIIE release and promoter escape. Thus, we replaced the TCCCAGT-3′AAGCTAGCTCACTCGAGCCGAGCAGACGTGCCTACGGACCGGTCGACT-3′ by the ACTGGGAATGCTGCGCTTCGTGATGTCGACCGGATGAGACGTGCCTACGGACCGGTCGACT-3′U. 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.

PICs in the closed conformation were assembled according to an in vitro transcription protocol with minor modifications. The assembly buffer contained 12 mM HEPES, pH 7.9, 0.12 mM EDTA, 12% glycerol, 8.25 mM MgCl2, 60 mM KCl, 1 mM DTT, 0.05% NP-40, 2.5 mM ZnCl2. The following purified proteins and nucleic acids were sequentially added into the assembly buffer: Pol II, TFII, TFII, DNA (template1−non-template1), TFII and TFII at final concentrations of 185 mM, 3.6 mM, 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 MgCl2, 50 mM KCl, 1 mM DTT, 0.05% NP-40, 2.5 mM ZnCl2) and an additional 5 min, the beads were washed 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 MgCl2, 50 mM KCl, 1 mM DTT, 0.05% NP-40, 1 unit−1 BSA-free Salt-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 TP−TFIIA−TFIIB−DNA−Pol II−TFIIH, the complex was added afterwards to the purified PIC at a final concentration of 100 nM.

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 EM sample preparation (either negative stain or cryo-plunging).

Electron microscopy. Negative stain samples of PIC complex and of free TFIIF were prepared using 400 mesh copper grid containing a constant carbon supporting layer. The grid was plasma cleaned for 10 s immediately before sample deposition using a Solarus plasma cleaner (Gatan) equipped with 75% argon/25% oxygen and sprayed with (Supelco) 3 μl of the diluted 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 kV at a nominal magnification of ~80,000 (1.37 Å per pixel). The data were collected using the Leginon data collection software 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 before sample deposition. An 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 blotted for 4s and immediately 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 kV using a 626 single-tilt cryo-transfer 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 1,500 and 3,200 images for each of the cryo data sets were collected using the M data set collection software.

Image processing. Negative stain data pre-processing was performed using the Appion processing environment. Particles were automatically selected from the micrographs using a difference of Gaussians (DoG) particle picker. The contrast transfer function (CTF) of each micrograph was estimated using both the ACE2 and CTFFind programs during data collection; the phases were flipped using CTFFind, and particle stacks were extracted using a box size of 256 × 256 pixels (except for both the TFIIF containing PIC complex and free TFIIF 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 4.5σ of the mean value. 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 (M-MSA-MRA) within the IMAGIC software. 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, 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 TFIIF containing PIC, which has a substantial extra mass with respect to Pol II, the negative stain reconstruction of the TBP−TFIIB−DNA−Pol II−TFIIF−TFIIE sample, after low-pass filtering to 60 Å, was instead used as the initial reference. The density reconstruction of the free TFIIF, the core density segmented from the negative stain TBP−TFIIA−TFIIB−DNA−Pol II−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 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 previously published TFIIF, 16,566 for as previously published TFIIF, 46,712 for as previously published TFIIF, and 13,023 for free TFIIF. 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–TFIIIF 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\(^63\).

51. Knuesel, M. T., Meyer, K. D., Bernecky, C. & Taatjes, D. J. The human CDK8 subcomplex is a molecular switch that controls Mediator coactivator function. Genes Dev. 23, 439–451 (2009).

52. Pal, M., Ponticelli, A. S. & Luse, D. S. The role of the transcription bubble and TFIIB in promoter clearance by RNA polymerase II. Mol. Cell 19, 101–110 (2005).

53. Voss, N. R., Yoshioka, C. K., Radermacher, M., Potter, C. S. & Carragher, B. DoG Picker and TiltPicker: software tools to facilitate particle selection in single particle electron microscopy. J. Struct. Biol. 166, 205–213 (2009).

54. Mallick, S. P., Carragher, B., Potter, C. S. & Kriegman, D. J. ACE: automated CTF estimation. Ultramicroscopy 104, 5–29 (2005).

55. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. J. Struct. Biol. 142, 334–347 (2003).

56. Sorzano, C. O. et al. XMIPP: a new generation of an open-source image processing package for electron microscopy. J. Struct. Biol. 148, 194–204 (2004).

57. van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R. & Schatz, M. A new generation of the IMAGIC image processing system. J. Struct. Biol. 116, 17–24 (1996).

58. Kostek, S. A. et al. Molecular architecture and conformational flexibility of human RNA polymerase II. Structure 14, 1691–1700 (2006).

59. Frank, J. et al. SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. J. Struct. Biol. 116, 190–199 (1996).

60. Heymann, J. B. & Belnap, D. M. Bsoft: image processing and molecular modeling for electron microscopy. J. Struct. Biol. 157, 3–18 (2007).

61. Lander, G. C. et al. Complete subunit architecture of the proteasome regulatory particle. Nature 482, 186–191 (2012).

62. van Dijk, M. & Bonvin, A. M. 3D-DART: a DNA structure modelling server. Nucleic Acids Res. 37, W235–W239 (2009).

63. Lu, X. J. & Olson, W. K. 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. Nature Protocols 3, 1213–1227 (2008).