Published in final edited form as:

Annu Rev Biophys. 2017 May 22; 46: 59–83. doi:10.1146/annurev-biophys-070816-033751.

Structural Insights into the Eukaryotic Transcription Initiation Machinery

Eva Nogales^{1,2,3}, Robert K. Louder⁴, and Yuan He⁵

¹Molecular and Cell Biology Department and QB3 Institute, University of California, Berkeley, California 94720-3220

²Howard Hughes Medical Institute, Berkeley, California 94720-3220

³Molecular Biophysics and Integrative Bioimaging Division, Lawrence Berkeley National Lab, California 94720-3220; enogales@lbl.gov

⁴Biophysics Graduate Group, University of California, Berkeley, California 94720-3220

⁵Department of Molecular Biosciences, Northwestern University, Evanston, Illinois 60208-3500

Abstract

Eukaryotic gene transcription requires the assembly at the promoter of a large preinitiation complex (PIC) that includes RNA polymerase II (Pol II) and the general transcription factors TFIID, TFIIA, TFIIB, TFIIF, TFIIIE, and TFIIF. The size and complexity of Pol II, TFIID, and TFIIF have precluded their reconstitution from heterologous systems, and purification relies on scarce endogenous sources. Together with their conformational flexibility and the transient nature of their interactions, these limitations had precluded structural characterization of the PIC. In the last few years, however, progress in cryo-electron microscopy (cryo-EM) has made possible the visualization, at increasingly better resolution, of large PIC assemblies in different functional states. These structures can now be interpreted in near-atomic detail and provide an exciting structural framework for past and future functional studies, giving us unique mechanistic insight into the complex process of transcription initiation.

Keywords

PIC; TFIID; TFIIF; TFIIIE; TFIIF; general transcription factors; cryo-EM

INTRODUCTION

Transcription of DNA genes into mRNA is critical to all forms of life. Not surprisingly, the active site in all RNA polymerases is significantly conserved across kingdoms (27). In contrast, the overall molecular complexity of the full polymerase ranges from the single polypeptide chain in the DNA-dependent RNA polymerases of bacteriophages and mitochondria to the 12-subunit eukaryotic RNA polymerase II (Pol II). Even more dramatic is the variation in the number and complexity of factors that associate with an RNA polymerase to initiate transcription. In bacteria, one out of a set of sigma factors helps the polymerase find the transcription start site (TSS) and open/stabilize the DNA transcription bubble. In eukaryotes, those two activities require a large set of general transcription factors

(GTFs) that have to assemble with Pol II near the TSS to form the preinitiation complex (PIC). Containing TFIIA, TFIIB, TFIID, TFIIF, TFIIE, TFIIH, and Pol II, the PIC ultimately includes close to 50 polypeptide chains (31, 54, 69). This molecular complexity has made difficult the systematic structural characterization of PIC assembly and function in transcriptional initiation in spite of decades of biochemical studies. The advent of cryo-electron microscopy (cryo-EM) as a practical structural biology technique in the last few years (59, 61) has resulted in unprecedented new insights into the molecular details of the eukaryotic transcription initiation process. Cryo-EM can overcome the challenges of sample scarcity, lack of stability, and conformational and compositional heterogeneity that thus far have hampered X-ray crystallographic studies of the PIC. Most importantly, it has the potential to produce ever more elaborate structures that include regulatory elements and thus to shed light onto the molecular mechanisms of gene expression regulation.

EARLY STUDIES OF LARGE PIC COMPONENTS

The essential character of eukaryotic gene transcription initiation has inspired many studies over the last three decades aimed at a mechanistic understanding of this molecular process. A richness of biochemical experiments led to the *in vitro* reconstitution of the transcription initiation process, defining the molecular components and the functional roles of the different initiation factors aiding Pol II to find the TSS, open the transcription bubble, and clear the promoter (68). Transcription initiation constitutes a major regulatory step in gene expression, and a myriad of cofactors, gene-specific activators and repressors, and chromatin modifying and remodeling complexes contribute to this regulation. Here, we concentrate on structural studies that concern the general machinery required for initiation of basal transcription (Figure 1). Our emphasis is on large components of the PIC that for many years remained highly challenging for structural studies. With ten or more subunits each, TFIID, TFIIH, and Pol II have proven refractory to efficient reconstitution and are generally obtained by purification from endogenous sources—most commonly, budding yeast or HeLa cells. Given their scarcity, especially for TFIID, the amounts and concentrations obtained are extremely limiting for structural characterization. The problem is exacerbated when trying to assemble a full PIC on DNA (60). There is a plethora of structures for individual protein subunits or domains within the PIC from X-ray crystallography and nuclear magnetic resonance studies (21). We refer to some of these only in the context of how they were used to help interpret low- to medium-resolution structures of larger PIC subcomplexes.

The X-ray crystallographic structure of yeast Pol II was a structural breakthrough (22,23), soon to be followed by a number of studies addressing different functional states of this complex enzyme (30, 81). Efforts toward a structural understanding of the transcription initiation process included the determination of structures of yeast Pol II bound to a nucleic acid scaffold and a fragment of TFIIB (12, 42). Together with the crystallographic structures of TATA binding protein (TBP)-TFIIB-DNA (78) and TBP-TFIIA-DNA (6), it became possible to propose a model for a minimal DNA-bound PIC containing TFIIA, TBP, TFIIB, and Pol II. This model was further expanded by Hahn and coworkers (15,35), using crosslinking data to include TFIIF and TFIIE. At the same time, EM had been used for the low-resolution visualization of TFIID and TFIIH. These two complexes are particularly challenging, as they are both difficult to produce in significant amounts and highly flexible.

TFIID is an ~1-MDa complex that is recruited to the promoter via specific recognition and binding of core promoter sequences. TFIID includes TBP, which is sufficient for basal transcription in vitro on TATA box-containing promoters (31, 69, 76), as well as 13 TBP-associated factors (TAFs) (2, 11) required for recognition of other core promoter sequences (10, 45, 75, 80), as well as for interaction with other components of the PIC (83), establishing TFIID as the primary core promoter recognition factor that nucleates PIC assembly. Furthermore, the TAFs of TFIID have been shown to interact with transcriptional activators and epigenetic marks, further suggesting that TFIID has a role in regulating PIC assembly and transcription initiation (2).

The first structural description of TFIID was our low-resolution, negative stain EM structure of the human complex (3), which showed TFIID to be shaped as a horseshoe of lobes A, B, and C surrounding a central cavity. The shape of TFIID was confirmed by the studies of Schultz and coworkers (7) of the budding yeast complex. These authors also carried out antibody labeling studies to position different TAFs and proposed that many existed in two copies located in distinct lobes of TFIID (47, 48). However, defining the structure of TFIID is complicated by the fact that the complex is highly flexible (33, 62). EM studies by Tjian and coworkers (49, 50) were early efforts to address the binding of activators to TFIID and the structural consequences of TAF isoforms.

TFIIH contains ten different subunits, including two ATPases—XPB and XPD—and the CDK kinase that forms part of a kinase module, or CAK. The ATPase activity of XPB is required for core promoter opening, and the phosphorylation of the C-terminal domain of Pol II by CDK is important for the transition of Pol II to the elongation phase (1, 40). The Schultz lab (72) was the first to produce a low-resolution structural description of TFIID in studies of the human complex. The authors proposed a model of the subunit distribution that was flawed owing to the fact that the CAK module was not actually present in the density map because of disorder. Two later EM studies, of both the yeast (29) and the human complexes (36), described the mobile character of the CAK subcomplex and positioned the two ATPases—XPB and XPD—within the better defined, core region of the complex. The yeast study by Kornberg and coworkers (29) made use of labeling strategies, including genetic tagging or genetic deletion of components. The study of the human complex by our lab positioned these two large proteins by docking homology models into the density ascribed to TFIID within a TBP-based PIC assembly (36). The TFIID shape and the subunit position of these two studies were highly consistent. More recent crosslinking and mass spectrometry studies and/or cryo-EM analysis have led to models of how other subunits are positioned within TFIID (37, 53, 58).

CRYO-EM STUDIES OF TFIID STRUCTURE, DYNAMICS, AND DNA BINDING

As mentioned above, early EM studies of TFIID defined lobes A, B, and C and indicated that the complex exhibits a large degree of conformational plasticity. This flexibility and its link to promoter binding was better characterized in later studies that used rigorous sorting, of both apo and DNA-bound-TFIID images, into distinct conformational states (18). The analysis showed the translocation of lobe A (comprising approximately one-third of the mass of the entire TFIID) from one side of the stable BC core to the other, traversing a

distance of over 100 Å. Quantification of lobe A positioning revealed a bimodal but continuous distribution between the two most abundant states, termed canonical and rearranged (Figure 2a). In the canonical state, lobe A is attached to lobe C, whereas the newly described rearranged state showed lobe A on the other side of the BC core, in close proximity to lobe B. The fact that complexes lacking lobe A were never observed indicates that lobe A stays tethered during the transition from one state to another. Of particular relevance is that the presence of both TFIIA and promoter DNA caused a shift toward the rearranged state (Figure 2a). The orthogonal tilt reconstruction (46) was used for ab initio generation of three-dimensional (3D) models of TFIID corresponding to the canonical and the rearranged states, which were then used as initial references in multi-model refinement of cryo-EM data for apo TFIID and TFIID in the presence of TFIIA and super core promoter. This analysis showed that TFIID binds promoter DNA only in the rearranged state (18) (Figure 2b).

Biochemical footprinting showed that TFIID interactions with the upstream and downstream sequences of the promoter DNA are distinct but coupled (18). TFIID interacted with promoters containing the downstream core promoter motifs (Inr, DPE, and M TE) independently of the presence of TFIIA, whereas protection of the TATA sequence by TFIID required the presence of TFIIA. Interestingly, even in the absence of a TATA sequence, TFIID was able to weakly protect the upstream region of the promoter in the presence of TFIIA and downstream sequences.

Recently, we significantly improved the resolution of the promoter-bound structure of TFIID through better handling of sample heterogeneity (52). We reduced compositional heterogeneity using a procedure for purifying promoter-bound complexes on the basis of our earlier work on TBP-based PICs (see the section Cryo-EM Studies of PIC Assembly and Architecture) (Figure 3a). We also used mild crosslinking (i.e., incubation with 0.01% glutaraldehyde for 5 min on ice) to help preserve the complexes during the harsh process of cryo-EM sample preparation. The most significant improvement came from the use of a K2 Summit direct detector (Gatan), which provided images with higher contrast and resolution that better allowed the sorting of conformational heterogeneity during image processing. For the latter, we used maximum-likelihood-based image classification and reconstruction in RELION(71) (see Figure 3b).

The improved structure of promoter-bound TFIID-TFIIA (EMD 3304) showed the overall organization described previously, but lobe A separated into a better defined, smaller lobe (lobe A1) that interacts stably with the promoter DNA, and a larger and more flexible lobe A2 (Figure 3b, *top*; and Figure 3c, *top*). By carrying out local classification and refinement within a mask that excluded flexible regions, we improved the reconstruction of the more stable parts of the complex (EMD 3305 and EM D 3306), ultimately reaching ~7-Å resolution for most of lobe C (Figure 3c, *middle* and *bottom*). Lobe A1 could be unambiguously ascribed to the TFIIA-TBP-TATA subcomplex by docking its crystal structure (6) into the density (Figure 4a, *left*). The localization of the TATA sequence allowed accurate positioning of the rest of the promoter sequences along the DNA density, confirming that all the downstream core promoter elements (Inr, M TE, DPE) are contacted by TFIID (Figure 4a, *center*). Further modeling of the TAF1-TAF7 subcomplex (PDB

4RGW) and TAF2 aminopeptidase-like domain into the density adjacent to the downstream promoter revealed that TAF1 and TAF2 mediate all the contacts with the downstream promoter (Figure 4a, *center* and *right*; and Figure 4b), using highly conserved positively charged interfaces (Figure 4b,c). TAF1 likely provides sequence-specific recognition of the conserved promoter sequences, and TAF2 likely provides ancillary interactions with the DNA. A majority of the downstream DNA contacts appeared to be made through the winged-helix (WH) domain of TAF1, in accordance with structural and biochemical findings (82) (Figure 4b, *left*).

Two copies of the TAF6 HEA T repeat domain (TAF6C; PDB 4ATG) could also be fitted into the density adjacent to TAF2, forming a nonsymmetrical homodimer that bridges the downstream promoter binding region of lobe C with the upstream promoter-interacting lobes B and A2 (Figure 4a, *center*). We proposed that an additional density for an alpha helix between the TAF6C dimer and the TAF2 aminopeptidase-like domain corresponds to a region of TAF8 that was shown to crosslink to TAF6 and be critical for the incorporation of TAF2 into TFIID (77) (Figure 4a, *center*). Although the asymmetric nature of the interface within the TAF6C dimer is curious, Berger and colleagues (5) have shown that addition of a single copy of TAF8 (and its histone-fold partner TAF10) to a dimeric TFIID subcomplex consisting of two copies each of TAF4, TAF5, TAF6, TAF9, and TAF12 converted the complex from a symmetric to an asymmetric structure. Thus, this symmetry breaking may correspond to an important transition during the assembly of TFIID, in which TAF8 incorporates the TFIID-specific subunits TAF1, TAF2, and TAF7 into the complex and activates it for core promoter binding.

The promoter-bound structure suggests that TFIID acts as a molecular ruler to place TBP on the upstream core promoter at a position that will be defined by the interactions of TAF1 and TAF2 with the downstream core promoter. A majority of eukaryotic promoters lack a consensus TATA sequence, underscoring the necessity of the sequence-specific interactions of TAFs with downstream core promoter elements for the accurate positioning of TBP on the upstream promoter, which ultimately determines the positioning of Pol II with respect to the TSS.

Some TAFs are likely to regulate loading of TBP onto the upstream core promoter. Existing structures show that TBP comigrates with lobe A during TFIID's transition from the canonical to the rearranged state (Figure 4d). In the canonical state, lobe A, and therefore TBP, is positioned adjacent to the TAF1-TAF7 module (Figure 4d, *left*). Previous structural and biochemical studies showed an interaction of the N terminus of TAF1 with both the concave and the convex surfaces of TBP (41). Although binding of the TAND to the concave surface inhibits TBP binding to DNA, TFIIA competes for binding to this TBP surface and effectively represses this inhibition. Thus, the canonical form of TFIID likely represents an autoinhibited state of the complex in which the TAND of TAF1 inhibits DNA binding by TBP.

The rearrangement of TFIID and its interaction with TFIIA are likely to be coupled events that play a role in the handoff of TBP to the upstream promoter DNA. Downstream promoter binding may also play a role in this transition, because the TAF1-TAF7 module appears to

undergo a significant reorganization upon binding the downstream promoter (Figure 4d, *center* and *right*). Such a mechanism of autoinhibition may be necessary to reduce nonspecific DNA binding by TBP in the absence of TFIIA and/or downstream core promoter sequences and thus prevent errant initiation. It may also present opportunities for regulating the promoter binding by TFIID through the strength of downstream core promoter sequences or the presence of factors that modulate the rearrangement of TFIID (19).

CRYO-EM STUDIES OF PIC ASSEMBLY AND ARCHITECTURE

In vitro reconstitution of the process of transcription initiation has led to a generally accepted model for a sequential assembly pathway (8). TFIID is first recruited to the promoter via specific recognition of core promoter sequences. TFIIA and TFIIB further stabilize the interaction between TFIID and promoter DNA. Next, Pol II, likely in association with TFIIF, adds to the growing PIC. Finally, TFIIH and TFIIE, which is required for DNA melting, are recruited to form the transcriptionally competent PIC (31, 69).

Structural characterization of the large PIC assembly is challenging and until recently had been limited to models based on chemical probing. Crystallographic structures of subsets of components, combined with these biochemical data, led to a number of partial structural models for the PIC, in either a closed (bound to duplex DNA) or an open (engaged with a transcription bubble) promoter conformation (17, 35, 42, 51). But direct visualization of the PIC was lacking. Our lab developed a reconstitution system that allowed the purification of active, homogenous human PIC assembly intermediates for cryo-EM studies that aimed to directly localize each G TF and track its effect on the PIC (36). We used a simplified PIC, in which TBP substituted for TFIID, that ultimately contained 31 polypeptides. Our promoter DNA contained TATA, BRE, and Inr core promoter elements and was immobilized on magnetic streptavidin beads through a biotin tag engineered at the 3' end of the non-template strand (similar to what is shown in Figure 3a). After incubation with the desired GTFs, stable complexes were released from the beads by restriction enzyme digestion. We devised a bootstrap approach that allowed us to identify, in a robust fashion, the position of each PIC component (Figure 5a, subpanels *i-iii*). The rationale was to first generate a structure for a PIC that included DNA-TBP-TFIIA-TFIIB-Pol II, for which there were crystallographic structures or homology models of all the components. This step was followed by purification and EM reconstruction of complexes with one additional component at a time, for which additional density in the EM map would most likely correspond to the newly added component. The final human PIC concentrations for cryo-EM studies were in the low hundreds of nanomolar and volumes realistically limited to 5–10 μ l. At these concentrations, mild crosslinking (i.e., incubation with 0.05% glutaraldehyde for ~5 min on ice) is useful for keeping the complexes together during the harsh process of cryo-EM sample preparation (60). The cryo-EM maps generated at the time (before direct detectors) had resolutions in the 12–25-Å range, allowing the docking of crystallographic structures available for some of the PIC components into the map.

The cryo-EM density map of the first PIC structure (Figure 5a, subpanel *i*) was fully accounted for by the existing crystallographic structures of TBP-TFIIB-DNA (78), TBP-

TFIIA-DNA (6), and yeast Pol II-TFIIB (42), with only some adjustment for the relative position of the two TFIIB cyclin folds. Only the upstream DNA, contacted in a sequence-specific manner by TBP and TFIIB, was visible, indicating that the rest was flexible and not engaged by Pol II. When the PIC was assembled including those factors plus TFIIF (a heterodimer of Rap30 and Rap74), the reconstruction showed additional density (Figure 5a, subpanel *ii*) that could be accounted for by the crystal structures of the WH domain of Rap30 (34) and the dimerization domain of Rap30-Rap74 (28). This TFIIF-containing PIC structure showed density for the full length of the DNA, indicating its stabilization. There were clear contacts of the DNA with the WH domain of Rap30 near the BREd and with the Rpb1 clamp head and Rpb5 components of Pol II, up- and downstream of the Inr, respectively. The presence of TFIIF also led to conformational changes that included the opening of the clamp (which allowed the engagement of DNA with this region of Rpb1), the slight bending of the downstream DNA, and the rotation toward Pol II of the upstream module formed by TBP, TFIIA, TFIIB, and the upstream DNA, as it engaged the WH domain of Rap30.

Purification and reconstruction of a complex that also included TFIIE showed extra density interacting with the Rap30 WH domain, the clamp head, and the stalk regions of Pol II (Figure 5a, subpanel *iii*). A model of three WH domains within TFIIE, interacting with the Pol II clamp head, had been proposed on the basis of crosslinking (35). That model approximately fit the elongated TFIIE density and ended by directly contacting the Rap30 WH domain. Thus, a continuous chain of four consecutive WH domains, linking the Pol II stalk region with the TBP-TFIIA-TFIIB-TFIIF-DNA subcomplex, topologically traps the DNA on the surface of Pol II. Finally, we purified and reconstructed a complex in the presence of TFIIH. Although excess TFIIE was needed in the previous step, indicating a low affinity of TFIIE for the growing PIC, such excess was no longer required in the presence of TFIIH, suggesting that TFIIE and TFIIH might be recruited cooperatively to the PIC, as previously proposed (20, 32, 69). The TFIIH-containing PIC showed a large extra density that could be assigned to TFIIH (not shown). Because this structure was initially obtained for a negatively stained sample, the DNA was not visible. The DNA position was extrapolated from the previous state, which turned out to be a good approximation based on later studies (37) (see Figure 5a, subpanel *v*). TFIIH was seen to contact the DNA right downstream of the Rpb5 contact, as well as the region of TFIIE proximal to the stalk (Figure 3a).

The self-consistency of this bootstrap strategy gave confidence to our experimental design, the cryo-EM reconstructions, and the molecular interpretation, allowing the unequivocal assignment of all the relative positions of proteins and DNA within the TBP-based human PIC. After the studies of the human PIC just described, Kornberg and colleagues published the structure of a TBP-based PIC from budding yeast (57) (Figure 5c, subpanel *i*), using an alternative assembly strategy in which all the yeast general factors (TFIIA to TFIIH) are proposed to form a stable subcomplex on DNA, which then engages Pol II. This fully assembled PIC was analyzed by cryo-EM to produce a structure that was interpreted as containing two main lobes, one corresponding to the Pol II (P lobe) and the other to the rest of the GTFs (G lobe). Interpretation of this reconstruction was based on docking of the Pol II structure into the P lobe and on a computational model of protein-protein interactions

from crosslinking and mass spectrometry data. Crystal structures of the available PIC components were then docked into the cryo-EM map, although little correspondence was apparent between the model and the EM density. The resulting yeast PIC model was not in agreement with the human one, which Kornberg and colleagues dismissed as erroneous.

Following the apparent discrepancies of the two PIC cryo-EM studies described, Cramer and colleagues (56) used crosslinking data and their own cryo-EM studies to describe yeast PIC models (lacking TFIIE and TFIIH) (64) (Figure 5b, subpanel *iii*) that more closely resembled that reported for the human complex. The initial disagreement was put to rest when an improved cryo-EM structure of the yeast PIC was reported by Kornberg and colleagues (58), showing the same subunit organization as initially reported for the human complex (Figure 5c, subpanel *ii*). In this study, one lobe corresponds to TFIIH and the other to Pol II and the rest of the GTFs.

More recently, additional cryo-EM structures of the yeast PIC have been obtained by the Cramer lab (63) (Figure 5b, subpanels *ii* and *iii*). These PICs were assembled following the stepwise assembly model. They lacked TFIIH, and the assembly required the presence of Mediator, although the density for this factor was not well resolved in the cryo-EM reconstructions and was ignored. The position of all the GTFs agreed generally with the positions observed in the human complex, with some small differences that may be of biological relevance (see the section Similarities and Differences Between the Human and Yeast PICs). Cramer and coworkers found that 75% of the particle images for the closed PIC complex actually corresponded to an open complex. The authors proposed that for their system, a transcription bubble can spontaneously form from duplex DNA and that the action of TFIIE traps it in such a state. This effect was not seen in the studies of the human PIC in our laboratory or the yeast studies by Murakami et al. Whether the difference is due to the promoter used, the presence of Mediator in the Cramer studies, or something else is not known. However, previous biochemical studies have also shown different requirements for promoter opening for different systems, with the most demanding requiring the ATPase action of TFIIH (39, 40).

INITIAL CRYO-EM STUDIES OF THE CLOSED TO OPEN PIC TRANSITION

The PIC remains stably associated during transcription initiation until Pol II undergoes promoter escape (69). Before that, the PIC needs to transition from a closed complex (CC) into an open complex (OC) in which the melted single-stranded DNA is inserted into the active site of Pol II.

To gain structural insight into the CC to OC transition, we generated a so-called functional mimic of the OC by replacing the segment of DNA containing the Inr with a 3'-tailed sequence based on previous studies of yeast Pol II (16) and matched the arrested position of Pol II on the template exactly to the TSS, thereby creating a Pol II-nucleic acid complex that included 5 nucleotides at the active site, while still containing upstream core promoter elements for assembling the rest of the PIC (36). The reconstruction of this OC mimic showed all the GTFs and the upstream DNA in almost identical positions to those in the CC. Although excess TFIIE had to be used to saturate the CC with this factor in the absence of

TFIIH, this was not the case for the open state mimic, indicating that TFIIE has a higher affinity for the PIC in the open state. The downstream DNA showed a conformation previously observed for elongating Pol II (23), indicating that the template strand was inserted into the active site. Comparison of the CC and OC structures showed a change in downstream DNA corresponding to a rotation on a plane as DNA translates, maintaining the point of contact with Rpb5 just downstream of the Inr. Unfortunately, the single-stranded segments were either invisible at the resolution of the reconstruction ($\sim 12\text{\AA}$) or not present (nontemplate strand). Another difference between the CC and OC reconstructions was the closing of the clamp domain in the OC to engage the DNA bubble, in a conformation that had been observed in the elongation state (30). As a consequence, the clamp domain appeared to complete a cycle, from closed to open and back to closed, throughout the process of PIC assembly and promoter opening, a cycle that had also been reported for the bacterial system (14).

HIGH-RESOLUTION CRYO-EM STUDIES OF THE PIC CLOSED COMPLEX

Cryo-EM has only recently become an efficient structural biology method capable of producing high-resolution structures. The leap in resolution is due to progress in detector technology (55) and the development of software tools with robust classification of image and reconstructions (71). Given the unique capacity of cryo-EM to deal with samples that are refractory to crystallization, these technical advances are resulting in an ever increasing number of atomic structures for molecules or assemblies that were considered structurally unreachable. In spite of these advances, certain samples remain challenging, even for cryo-EM. Large eukaryotic transcription complexes are an example, as they suffer from scarcity of the sample, poor stability of the complexes, compositional heterogeneity, and, most often, severe intrinsic flexibility (60). Efforts to overcome these limitations, together with the new technical advances in the field, have now resulted in recent high-resolution cryo-EM structures of PICs in different functional states [CC, OC, and initially transcribing complex (ITC)], for both the human and budding yeast systems by the Nogales and Cramer labs, respectively (37, 63). These structures, leading to atomic models for most of the PIC components, are contributing significant new mechanistic information toward our understanding of the functional molecular transitions that define the transcription initiation process.

Structures of yeast and human PIC in the CC state are still the most limited in resolution, although the use of direct detectors have made it possible for three labs to reach subnanometer resolution, at least for the so-called core PIC lacking TFIIH (Figure 5a, subpanel *iv*; Figure 5b, subpanel *i*; and Figure 5c, subpanel *iii*). In the case of the studies of the human and yeast TFIIH-containing PIC (37, 58), the limiting factor was the mobility of TFIIH. For the human complex, the range and direction of motion were described for all states studied—CC, OC, and ITC (37). To improve the resolution of the core PIC in the CC [comprising TBP, TFIIA, TFIIB, TFIIF, TFIIE, TFIIIS, and Pol II (and Sub1 for the yeast structure)], alignment of the particle images was carried out masking out the TFIIH density within the running reference structure during refinement. In the case of the study by Cramer and coworkers (63), the limitation had to do with the fact that most of the particle images

corresponded to a spontaneously generated OC state from the duplex promoter DNA, resulting in limited amount of data useful for the CC reconstruction.

Comparison of all the CC structures shows clear correspondence for the relative positions of all the GTFs with respect to Pol II (Figure 6 shows the model for the human CC, with GTFs labeled). However, detailed analysis reveals differences that may reflect slight functional divergence between the human and yeast systems, ultimately concerning the opening of the promoter and the positioning of the TSS. One of the most obvious differences between the yeast and human CC structures is that although there are no contacts between DNA and Pol II for the yeast structures, the human Pol II makes contacts around the Inr sequence via the clamp head of Rpb1 and via Rpb5. As a result, the path of the DNA is clearly distinct (Figure 7a). Most likely related to DNA engagement, the clamp head of the yeast Pol II is closed, resembling that of an elongating polymerase, whereas that in the human Pol II is open. Recall that in the previous studies by He et al. (36), this opening occurred upon addition of TFIIF, leading to the interaction and stabilization of the DNA. In the Murakami structure, the DNA is engaged further downstream by TFIH, as it is for the He et al. structure (37, 52). Because TFIH was missing from the study by Plaschka et al., part of the DNA is missing in the reconstruction because of flexibility (68).

There are a number of differences in the position/conformation of some of the GTFs. The most significant concerns the position of two adjacent WH domains in TFIIF and TFIIE (Figure 8a-c). The WH domain of Rap30 (Tfg2 in yeast) within TFIIF is displaced by approximately 15 Å between the human and yeast structures and contains an additional helix because of a significant insertion for the yeast sequence (Figure 8a, *dashed line*). By contrast, the dimerization domain within TFIIF [containing part of Rap30 and Rap74 (Tgfl in yeast)], which contacts the jaws regions of Rpb1 and Rpb9 and the Rpb2 external 2, is in a very similar position for both systems (Figure 8b). The arm domain of Tgfl, however, is stably bound to the protrusion domain, whereas its counterpart in the human Rap74 could not be clearly seen and appears to be mostly disordered.

The WH1 domain of TFIIEβ (Tfa2 in yeast) that interacts with the Rap30–Tfg2 WH domain and in particular with the additional helix of this domain in yeast is, accordingly, also in a different position between the human and yeast systems (Figure 8c). Because of the relatively low resolution of this region in all the reported structures, position is better defined than the details of the homology model docked into the map; but notice that there were four independent human structures, so there is redundancy for the human system. The WH2 domain of TFIIEβ–Tfa2 has a slightly different position in the two systems. Interestingly, other regions of TFIIE (farther away from the DNA and TFIIF) are more similar between the two systems (Figure 8c). The helix-turn-helix and WH domain of TFIIEα–Tfal (with the exception of the E-wing) are basically superimposable in the Plaschka and He studies (37, 63). This is even the case for the TFIIE α–Tfal zinc ribbon (E-ribbon) that interacts with the Rpb4–Rpb7 stalk, although the whole stalk module is in a slightly different position with respect to the rest of Pol II in the two systems. When we aligned the modeled regions of TFIH in the study of the human complex by He et al. and the yeast study by Murakami et al. (Figure 8e), the downstream DNA that is engaged by XPB is almost superimposable, as

is the TATA box engaged by TBP 200 Å away. Between these two sites, the DNA is kinked, with a sharper bend in the yeast structure.

HIGH-RESOLUTION CRYO-EM STUDIES OF THE PIC OPEN COMPLEX

The Cramer and Nogales labs have now reported high-resolution ($\sim 3.5\text{--}7\text{\AA}$) structures of the yeast and human OC states, respectively, in the absence of TFIIH (37, 63) (Figure 5a, subpanel *iv* and Figure 5b, subpanel *ii*). These core OC structures are again similar for human and yeast, with the differences corresponding to the position of some of the domains in TFIIF and TFIIE remaining. In both cases, the authors used a mismatch bubble to mimic the transcription bubble when biochemically generating an OC state for cryo-EM study. But the yeast study by Cramer and colleagues also reported an OC state generated spontaneously and indicated that it was practically indistinguishable from the OC state purposely designed.

The one marked difference between the structures of the human and yeast OCs is that although single-stranded DNA is not visible in the yeast structure, a full bubble is visible in the human OC (Figure 7b). Additionally, the B-reader and B-linker domains of TFIIB appeared poorly defined as a result of disorder in the yeast structure, but the human structure shows that the TFIIB linker and reader regions (which were not visible in the human CC structure) have become stabilized with respect to the CC by interaction with the single-stranded DNA. Finally, the E-wing of TFIIE is in clear contact with the upstream end of the bubble in the yeast structure, whereas no density near DNA can be ascribed to human TFIIE (Figure 7b). This difference may be the reason why the human PIC does not spontaneously open the promoter. The higher stability of the DNA bubble and key regions of TFIIB in the human system with respect to the yeast system may reflect the distinct positions of the TSS with respect to the TATA box for the two systems. Although this position is more defined for the human PIC (9), there is more variability in yeast, which engages in a scanning process following bubble opening before transcription starts (73). This scanning process will give rise to a variety of nucleic acid states for the bubble and thus to blurring in the cryo-EM map.

For the human system, in addition to the high-resolution structure of the core OC, there was a lower-resolution structure that included TFIIH (Figure 5a, subpanel *vii*). Thus, comparison of this part of the structure indicated how the XPB helicase and the DNA that it interacts with move from the CC to the OC states (Figure 9). The transition involves a slight rotation of TFIIH approximately in the plane containing the DNA as it bends into the open bubble (Figure 9d).

HIGH-RESOLUTION CRYO-EM STUDIES OF THE PIC INITIALLY TRANSCRIBING COMPLEX

He et al. have obtained both a structure of an ITC that includes TFIIH at lower resolution (between 6 and 20 Å) (Figure 5a, subpanel *ix*) and a high-resolution and atomic model for the core PIC without TFIIH (Figure 5a, subpanel *viii*). The cryo-EM map of the human core ITC shows clear density for the RNA and for all single-stranded DNA, except for approximately 3 nucleotides of the nontranscribed strand at a site where we believe

scrunching occurs. This site faces the cavity sandwiched between the protrusion and the lobe region of Pol II, very close to where the arm domain of TFIIF was proposed to reside in the Cramer structure (63). There is very little difference between the human O C and ITC states, except for the larger DNA bubble and the presence of RNA. The position of all GTFs is unchanged, including that of TFIIH, which already in the O C state is positioned to be able to continue translocating the downstream DNA into the Pol II active site, in excellent agreement with its role of aiding in transcription initiation and eventually in promoter escape (26). A minimal yeast ITC containing TBP, TFIIB, TFIIF, and Pol II had been previously described by Cramer and colleagues (Figure 5b, subpanel *iii*). This yeast structure did contain most of the transcription bubble and RNA. Comparison with the human ITC shows a small rotation of the TBP-TFIIB-TATA lobe and the stalk of Pol II, which was also present when comparing other states. The Tfg2 WH domain was clearly mobile, resulting in an almost invisible density in the corresponding region, perhaps because of the lack of association with TFIIE. Interestingly, the region of apparent DNA disorder in the human ITC, which was interpreted as a likely location of scrunching, is a region well defined in the yeast ITC. This difference may be due to the specific DNA constructs used for the two studies.

SIMILARITIES AND DIFFERENCES BETWEEN THE HUMAN AND YEAST PICs

Comparison of all the existing cryo-EM structures of human and yeast PICs in different states leads to an overall understanding of similarities and differences between the two systems. It is clear that the general position of all GTFs is highly conserved between the two systems, as expected from the essential character of the transcription initiation process. However, but not surprisingly either, previously described small functional disparities between these two well-studied systems are reflected in small but nevertheless significant structural differences that illuminate the basis of such differences in the context of an overall well-conserved process. The functional differences relate to the capacity of the yeast PIC to open certain promoters without the requirement of TFIIH, and the lack of a scanning process in search of the TSS in the human system, where the position of the TSS is precisely defined by the position of the TATA box. The structural differences that may likely relate to the functional divergence are more significant when it comes to the position and stability of the DNA. In the human system, the closed promoter is engaged by Pol II around the Inr core promoter motif, which includes the TSS, in addition to the engagement that is also seen in yeast by the upstream factors TBP, TFIIB, and TFIIF (i.e., the WH domain of Rap30) and the XPB ATPase within TFIIH on the downstream end. In contrast, contacts between the duplex DNA and Pol II have not been seen for any of the yeast CC structures reported. Likely influencing the absence of such contacts is the fact that several structural motifs within TFIIF, TFIIE, and TFIIB are in clearly distinct positions (or have different degrees of structural stability) in the human and yeast systems. These obvious differences lead to slight repositioning of the upstream GTFs and to an open state of the clamp head in the human system that is not seen in any of the available cryo-EM structures from yeast, and that very likely allows the human Pol II to engage duplex DNA. The differences in G TF positions just mentioned are also visible in the OC state for the two systems. Additionally, significantly

more stable density is apparent in the human OC complex that accounts for the full, stably engaged transcription bubble within the Pol II cleft, as well as the stable engagement of regions of TFIIB with the single-stranded DNA. Neither of these two elements can be seen in the yeast OC, likely because of inherent variability in the scanning process.

From the PIC structures recently obtained, different models have been proposed for the transitions between the CC, OC, and ITC states. Given that only the studies of the human PIC (a) visualize the full length of the DNA in all states and (b) have analyzed the three states with a full set of GTFs, we have concentrated this review on the human structures to describe these transitions. On the basis of similarities and differences we have described for the structures of two systems, we expect that small differences may exist between yeast and human but within a framework that must be generally conserved. In particular, the data available concerning the position of XPB (Ssl2 in yeast) (Figure 8e) within the PIC agree very well with the proposals of a translocase activity for this TFIIH protein (35, 36, 40) and with the function of TFIIH, not only in the opening of the transcription bubble but in promoter escape, as the location of XPB in the OC and ITC human complexes is identical and shows how translocation of DNA will continue during the initial synthesis of RNA, facilitating the feeding of the transcribed strand into the active site as this early transcription proceeds (Figure 9d).

STRUCTURAL STUDIES OF LARGER PIC COMPLEXES

In the studies of the PIC described so far, the largest assembly included TBP, TFIIA, TFIIB, TFIIF, TFIIE, and TFIIH (as well as TFIIS or Sub1) (37, 58). In these studies, TBP substituted for the full TFIID. Although there is no experimental structure of a TFIID-based PIC yet, we generated a so-called synthetic model based on the superposition of the common elements between the TBP-based PIC and TFIID-TFIIA-DNA structures (i.e., TBP, TFIIA, and TATA box DNA) (Figure 10a). This superposition shows an overall complementarity in shape that allows for a number of contacts with minimal but interesting clashes (52). Lobe B, residing on the upstream end of the PIC, is in a perfect location to make contact with either TFIIF or TFIIE, both of which have been shown biochemically to interact directly with TFIID (25, 38, 70). The presence of such interactions could implicate TAFs in having a role in PIC assembly beyond the initial binding to promoter DNA. Notably, the promoter-binding site of the Rap30 WH domain overlaps the -16 binding site of TFIID, suggesting that a reorganization occurs in this region upon TFIIF recruitment.

Pol II sits between the upstream and downstream promoter contacts of TFIID, with its Rpb5 subunit making minor clashes with the promoter-bound TAF1 subunit of TFIID, suggesting further structural reorganization occurs in this part of the complex following the TFIIF-induced promoter engagement by Pol II. Indeed, time-course footprinting analysis has shown that a structural isomerization occurs following the addition of Pol II-TFIIB-TFIIF to the promoter-bound TFIID-TFIIA complex, largely characterized by the release of TFIID-dependent protection of downstream promoter DNA with a concomitant increase in the protection of the promoter upstream of this region as it engages Pol II (84). Together, these results point to the release of downstream promoter contacts by the TAF1 and TAF2 subunits of TFIID following the recruitment of TFIIB-Pol II-TFIIF, which we speculate is related to

the subsequent need of TFIIH to bind the downstream DNA and melt the promoter using its helicase activity, as well as the need of Pol II to eventually transcribe through the DNA downstream of the initial transcription bubble (Figure 10b).

Recently, a tin(IV) oxochloride cluster was found to inhibit transcription initiation by preventing Pol II recruitment through the inhibition of this isomerization, further indicating that this rearrangement is a necessary step in PIC assembly. Interestingly, addition of the metal cluster to PICs that had already undergone one round of transcription initiation had no effect on the rate of reinitiation from these complexes, suggesting that the isomerization is irreversible and occurs only during de novo initiation (85). Thus, the current data suggest that the sequence-specific promoter interactions between TFIID and the downstream core promoter are important for initial PIC assembly and the first round of transcription initiation but may be dispensable for reinitiation.

Mediator is a large complex that interacts with the PIC and with gene-specific activators to regulate transcription. A number of studies have elucidated the crystal structure of different regions of the yeast complex (see 44 and 65 for recent reviews), including those of the Mediator head subcomplex (43, 66). But, because of its size and complexity, the full Mediator has proved difficult to crystallize and has been traditionally pursued by EM, both for the human and yeast systems. Early studies in negative stain defined the overall architecture of these complexes (24, 74), whereas more recent studies of frozen-hydrated samples started to define functional regions (79). There have also been EM efforts to visualize the interaction of Mediator with Pol II (4, 13). For the yeast system, recent cryo-EM studies have shed light on the interaction of Mediator with either a partial (64) or a full TBP-based PIC (67). These studies have identified the major regions of contacts between Mediator and Pol II and the GTFs, shedding light on Mediator's role in PIC stabilization and recruitment. Comparison of these structures with those of the synthetic TFIID-based PIC indicate that Mediator and TFIID sit on opposite sides of this large supracomplex, surrounding Pol II (52) (Figure 10c).

CONCLUDING REMARKS

Progress in biochemical reconstitution of the transcription PIC for structural analysis and advantages in EM detectors and image analysis software are leading to a marked progress in the mechanistic characterization of the complex transcription initiation process in eukaryotes via visualization of these transient complexes, in different functional states, using cryo-EM. The advantages of this technique, which can handle large assemblies, small amounts of sample, and, to a large extent, conformational and compositional heterogeneity, are making possible the analysis of complexes, such as the PIC, that have previously proven to be refractory to structural analysis. Parallel progress with the human and budding yeast systems is illuminating both the overall conservation of this machinery and the subtle structural differences that reflect small functional disparities previously described biochemically. Both similarities and differences are a rich source of mechanistic information to better understand the molecular transitions required to initiate transcription. The molecular assemblies that are required for regulated transcription initiation at a eukaryotic promoter can ultimately include cofactors, activators or repressors, and the right chromatin context. Although the

biochemical complexity will increase the compositional and conformational heterogeneity of these samples, these large and challenging transcription complexes may soon be within reach. The effort will require the use of innovative image analysis approaches, implemented in large data sets with improved image contrast. The resulting structures hold promise to be bursting with biological insight on the regulation of this critical step in gene expression.

Acknowledgments

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

1. Akoulitchiev S, Mäkelä TP, Weinberg RA, Reinberg D. 1995 Requirement for TFIIF kinase activity in transcription by RNA polymerase II. *Nature* 377:557–60 [PubMed: 7566158]
2. Albright SR, Tjian R. 2000 TAPs revisited: More data reveal new twists and confirm old ideas. *Gene* 242:1–13 [PubMed: 10721692]
3. Andel F, III, Ladurner AG, Inouye C, Tjian R, Nogales E. 1999 Three-dimensional structure of the human TFIID-IIA-IIB complex. *Science* 286:2153–56 [PubMed: 10591646]
4. Bernecky C, Grob P, Ebmeier CC, Nogales E, Taatjes DJ. 2011 Molecular architecture of the human Mediator-RNA polymerase II-TFIIF assembly. *PLOS Biol.* 9:e1000603 [PubMed: 21468301]
5. Bieniossek C, Papai G, Schaffitzel C, Garzoni F, Chaillet M, et al. 2013 The architecture of human general transcription factor TFIID core complex. *Nature* 493:699–702 [PubMed: 23292512]
6. Bleichenbacher M, Tan S, Richmond TJ. 2003 Novel interactions between the components of human and yeast TFIIA/TBP/DNA complexes. *J. Mol. Biol.* 332:783–93 [PubMed: 12972251]
7. Brand M, Leurent C, Mallouh V, Tora L, Schultz P. 1999 Three-dimensional structures of the TAFii-containing complexes TFIID and TFTC. *Science* 286:2151–53 [PubMed: 10591645]
8. Buratowski S, Hahn S, Guarente L, Sharp PA. 1989 Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell* 56:549–61 [PubMed: 2917366]
9. Buratowski S, Sotia M, Greenblatt J, Sharp PA. 1991 RNA polymerase II-associated proteins are required for a DNA conformation change in the transcription initiation complex. *PNAS* 88:7509–13 [PubMed: 1881889]
10. Burke TW, Kadonaga JT. 1997 The downstream core promoter element, DPE, is conserved from *Drosophila* to humans and is recognized by TAFii60 of *Drosophila*. *Genes Dev.* 11:3020–31 [PubMed: 9367984]
11. Burley SK, Roeder RG. 1996 Biochemistry and structural biology of transcription factor IID (TFIID). *Annu. Rev. Biochem.* 65:769–99 [PubMed: 8811195]
12. Bushnell DA, Westover KD, Davis RE, Kornberg RD. 2004 Structural basis of transcription: an RNA polymerase II-TFIIB cocrystal at 4.5 angstroms. *Science* 303:983–88 [PubMed: 14963322]
13. Cai G, Chaban YL, Imasaki T, Kovacs JA, Calero G, et al. 2012 Interaction of the mediator head module with RNA polymerase II. *Structure* 20:899–910 [PubMed: 22579255]
14. Chakraborty A, Wang D, Ebright YW, Korlann Y, Kortkhonja E, et al. 2012 Opening and closing of the bacterial RNA polymerase clamp. *Science* 337:591–95 [PubMed: 22859489]
15. Chen H-T, Warfield L, Hahn S. 2007 The positions of TFIIF and TFIIE in the RNA polymerase II transcription preinitiation complex. *Nat. Struct. Mol. Biol.* 14:696–703 [PubMed: 17632521]
16. Cheung AC, Cramer P. 2011 Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature* 471:249–53 [PubMed: 21346759]
17. Chung W-H, Craighead JL, Chang W-H, Ezeokkonkwo C, Bareket-Samish A, et al. 2003 RNA polymerase II/TFIIF structure and conserved organization of the initiation complex. *Mol. Cell* 12:1003–13 [PubMed: 14580350]

18. Cianfrocco MA, Kassavetis GA, Grob P, Fang J, Juven-Gershon T, et al. 2013 Human TFIID binds to core promoter DNA in a reorganized structural state. *Cell* 152:120–31 [PubMed: 23332750]
19. Cianfrocco MA, Nogales E. 2013 Regulatory interplay between TFIID's conformational transitions and its modular interaction with core promoter DNA. *Transcription* 4:120–26 [PubMed: 23863784]
20. Conaway RC, Conaway JW. 1993 General initiation factors for RNA polymerase II. *Annu. Rev. Biochem* 62:161–90 [PubMed: 8352588]
21. Cramer P 2014 A tale of chromatin and transcription in 100 structures. *Cell* 159:985–94 [PubMed: 25416940]
22. Cramer P, Bushnell DA, Fu J, Gnatt AL, Maier-Davis B, et al. 2000 Architecture of RNA polymerase II and implications for the transcription mechanism. *Science* 288:640–49 [PubMed: 10784442]
23. Cramer P, Bushnell DA, Kornberg RD. 2001 Structural basis of transcription: RNA polymerase II at 2.8 Å resolution. *Science* 292:1863–76 [PubMed: 11313498]
24. Dotson MR, Yuan CX, Roeder RG, Myers LC, Gustafsson CM, et al. 2000 Structural organization of yeast and mammalian mediator complexes. *PNAS* 97:14307–10 [PubMed: 11114191]
25. Dubrovskaya V, Lavigne AC, Davidson I, Acker J, Staub A, Tora L. 1996 Distinct domains of hTAFII100 are required for functional interaction with transcription factor TFIIF beta (RAP30) and incorporation into the TFIID complex. *EMBO J.* 15:3702–12 [PubMed: 8758937]
26. Dvir A, Conaway JW, Conaway RC. 2001 Mechanism of transcription initiation and promoter escape by RNA polymerase II. *Curr. Opin. Genet. Dev* 11:209–14 [PubMed: 11250146]
27. Ebright RH. 2000 RNA polymerase: structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. *J. Mol. Biol* 304:687–98 [PubMed: 11124018]
28. Gaiser F, Tan S, Richmond TJ. 2000 Novel dimerization fold of RAP30/RAP74 in human TFIIF at 1.7 Å resolution. *J. Mol. Biol* 302:1119–27 [PubMed: 11183778]
29. Gibbons BJ, Brignole EJ, Azubel M, Murakami K, Voss NR, et al. 2012 Subunit architecture of general transcription factor TFIIF. *PNAS* 109:1949–54 [PubMed: 22308316]
30. Gnatt AL, Cramer P, Fu J, Bushnell DA, Kornberg RD. 2001 Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science* 292:1876–82 [PubMed: 11313499]
31. Goodrich JA, Cutler G, Tjian R. 1996 Contacts in context: promoter specificity and macromolecular interactions in transcription. *Cell* 84:825–30 [PubMed: 8601306]
32. Goodrich JA, Tjian R. 1994 Transcription factors IIE and IIH and ATP hydrolysis direct promoter clearance by RNA polymerase II. *Cell* 77:145–56 [PubMed: 8156590]
33. Grob P, Cruse MJ, Inouye C, Peris M, Penczek PA, et al. 2006 Cryo-electron microscopy studies of human TFIID: conformational breathing in the integration of gene regulatory cues. *Structure* 14:511–20 [PubMed: 16531235]
34. Groft CM, Uljon SN, Wang R, Werner MH. 1998 Structural homology between the Rap30 DNA-binding domain and linker histone H5: implications for preinitiation complex assembly. *PNAS* 95:9117–22 [PubMed: 9689043]
35. Grunberg S, Warfield L, Hahn S. 2012 Architecture of the RNA polymerase II preinitiation complex and mechanism of ATP-dependent promoter opening. *Nat. Struct. Mol. Biol* 19:788–96 [PubMed: 22751016]
36. He Y, Fang J, Taatjes DJ, Nogales E. 2013 Structural visualization of key steps in human transcription initiation. *Nature* 495:481–86 [PubMed: 23446344]
37. He Y, Yan C, Fang J, Inouye C, Tjian R, et al. 2016 Near-atomic resolution visualization of human transcription promoter opening. *Nature* 533:359–65 [PubMed: 27193682]
38. Hisatake K, Ohta T, Takada R, Guermah M, Horikoshi M, et al. 1995 Evolutionary conservation of human TATA-binding-peptide-associated factors TAFII31 and TAFII80 and interactions of TAFII80 with other TAFs and with general transcription factors. *PNAS* 92:8195–99 [PubMed: 7667268]
39. Holstege FC, Tantin D, Carey M, van der Vliet PC, Timmers HT. 1995 The requirement for the basal transcription factor IIE is determined by the helical stability of promoter DNA. *EMBO J.* 14:810–19 [PubMed: 7882984]

40. Kim T-K, Ebright RH, Reinberg D. 2000 Mechanism of ATP-dependent promoter melting by transcription factor IIH. *Science* 288:1418–22 [PubMed: 10827951]
41. Kokubo T, Swanson MJ, Nishikawa J-I, Hinnebusch AG, Nakatani Y. 1998 The yeast TAF145 inhibitory domain and TFIIA competitively bind to TATA-binding protein. *Mol. Cell. Biol* 18:1003–12 [PubMed: 9447997]
42. Kostrewa D, Zeller ME, Armache K-J, Seizl M, Leike K, et al. 2009 RNA polymerase II-TFIIB structure and mechanism of transcription initiation. *Nature* 462:323–30 [PubMed: 19820686]
43. Lariviere L, Plaschka C, Seizl M, Wenzek L, Kurth F, Cramer P. 2012 Structure of the Mediator head module. *Nature* 492:448–51 [PubMed: 23123849]
44. Lariviere L, Seizl M, Cramer P. 2012 A structural perspective on Mediator function. *Curr. Opin. Cell Biol* 24:305–13 [PubMed: 22341791]
45. Lee D-H, Gershenson N, Gupta M, Ioshikhes IP, Reinberg D, Lewis BA. 2005 Functional characterization of core promoter elements: the downstream core element is recognized by TAF1. *Mol. Cell. Biol* 25:9674–86 [PubMed: 16227614]
46. Leschziner AE, Nogales E. 2006 The orthogonal tilt reconstruction method: an approach to generating single-class volumes with no missing cone for ab initio reconstruction of asymmetric particles. *J. Struct. Biol* 153:284–99 [PubMed: 16431136]
47. Leurent C, Sanders SL, Demeny MA, Garbett KA, Ruhlmann C, et al. 2004 Mapping key functional sites within yeast TFIID. *EMBO J.* 23:719–27 [PubMed: 14765106]
48. Leurent C, Sanders SL, Ruhlmann C, Mallouh V, Weil PA, et al. 2002 Mapping histone fold TAFs within yeast TFIID. *EMBO J.* 21:3424–33 [PubMed: 12093743]
49. Liu W-L, Coleman RA, Grob P, King DS, Florens L, et al. 2008 Structural changes in TAF4b-TFIID correlate with promoter selectivity. *Mol. Cell* 29:81–91 [PubMed: 18206971]
50. Liu W-L, Coleman RA, Ma E, Grob P, Yang JL, et al. 2009 Structures of three distinct activator-TFIID complexes. *Genes Dev.* 23:1510–21 [PubMed: 19571180]
51. Liu X, Bushnell DA, Wang, Calero, Kornberg. 2010 Structure of an RNA polymerase II-TFIIB complex and the transcription initiation mechanism. *Science* 327:206–9 [PubMed: 19965383]
52. Louder RK, He Y, Lopez-Blanco JR, Fang J, Chacon P, Nogales E. 2016 Structure of promoter-bound TFIID and model of human pre-initiation complex assembly. *Nature* 531:604–9 [PubMed: 27007846]
53. Luo J, Cimermanic P, Viswanath S, Ebmeier CC, Kim B, et al. 2015 Architecture of the human and yeast general transcription and DNA repair factor TFIH. *Mol. Cell* 59:794–806 [PubMed: 26340423]
54. Matsui T, Segall J, Weil PA, Roeder RG. 1980 Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. *J. Biol. Chem* 255:11992–96 [PubMed: 7440580]
55. McMullan G, Faruqi AR, Henderson R. 2016 Direct electron detectors. *Methods Enzymol.* 579:1–17 [PubMed: 27572721]
56. Mühlbacher W, Sainsbury S, Hemann M, Hantsche M, Neyer S, et al. 2014 Conserved architecture of the core RNA polymerase II initiation complex. *Nat. Commun* 5:4310 [PubMed: 25007739]
57. Murakami K, Elmlund H, Kalisman N, Bushnell DA, Adams CM, et al. 2013 Architecture of an RNA polymerase II transcription pre-initiation complex. *Science* 342:1238724 [PubMed: 24072820]
58. Murakami K, Tsai KL, Kalisman N, Bushnell DA, Asturias FJ, Kornberg RD. 2015 Structure of an RNA polymerase II preinitiation complex. *PNAS* 112:13543–48 [PubMed: 26483468]
59. Nogales E 2016 The development of cryo-EM into a mainstream structural biology technique. *Nat. Methods* 13:24–27 [PubMed: 27110629]
60. Nogales E, Louder RK, He Y. 2016 Cryo-EM in the study of challenging systems: the human transcription pre-initiation complex. *Curr. Opin. Struct. Biol* 40:120–27 [PubMed: 27689812]
61. Nogales E, Scheres SH. 2015 Cryo-EM: a unique tool for the visualization of macromolecular complexity. *Mol. Cell* 58:677–89 [PubMed: 26000851]
62. Papai G, Tripathi MK, Ruhlmann C, Werten S, Crucifix C, et al. 2009 Mapping the initiator binding Taf2 subunit in the structure of hydrated yeast TFIID. *Structure* 17:363–73 [PubMed: 19278651]

63. Plaschka C, Hantsche M, Dienemann C, Burzinski C, Plitzko J, Cramer P. 2016 Transcription initiation complex structures elucidate DNA opening. *Nature* 533:353–58 [PubMed: 27193681]
64. Plaschka C, Lariviere L, Wenzek L, Seizl M, Hemann M, et al. 2015 Architecture of the RNA polymerase II-Mediator core initiation complex. *Nature* 518:376–80 [PubMed: 25652824]
65. Poss ZC, Ebmeier CC, Taatjes DJ. 2013 The Mediator complex and transcription regulation. *Crit. Rev. Biochem. Mol. Biol* 48:575–608 [PubMed: 24088064]
66. Robinson PJ, Bushnell DA, Trnka MJ, Burlingame AL, Kornberg RD. 2012 Structure of the Mediator Head module bound to the carboxy-terminal domain of RNA polymerase II. *PNAS* 109:17931–35 [PubMed: 23071300]
67. Robinson PJ, Trnka MJ, Bushnell DA, Davis RE, Mattei P-J, et al. 2016 Structure of a complete Mediator-RNA polymerase II pre-initiation complex. *Cell* 166:1411–22e16 [PubMed: 27610567]
68. Roeder RG. 1996 Nuclear RNA polymerases: role of general initiation factors and cofactors in eukaryotic transcription. *Methods Enzymol.* 273:165–71 [PubMed: 8791609]
69. Roeder RG. 1996 The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem. Sci* 21:327–35 [PubMed: 8870495]
70. Ruppert S, Tjian R. 1995 Human TAFII250 interacts with RAP74: implications for RNA polymerase II initiation. *Genes Dev.* 9:2747–55 [PubMed: 7590250]
71. Scheres SH. 2016 Processing of structurally heterogeneous cryo-EM data in RELION. *Methods Enzymol.* 579:125–57 [PubMed: 27572726]
72. Schultz P, Fribourg S, Poterszman A, Mallouh V, Moras D, Egly JM. 2000 Molecular structure of human TFIIF. *Cell* 102:599–607 [PubMed: 11007478]
73. Struhl K. 1987 Promoters, activator proteins, and the mechanism of transcriptional initiation in yeast. *Cell* 49:295–97 [PubMed: 2882858]
74. Taatjes DJ, Näär AM, Andel F, III, Nogales E, Tjian R. 2002 Structure, function, and activator-induced conformations of the CRSP coactivator. *Science* 295:1058–62 [PubMed: 11834832]
75. Theisen JW, Lim CY, Kadonaga JT. 2010 Three key subregions contribute to the function of the downstream RNA polymerase II core promoter. *Mol. Cell. Biol* 30:3471–79 [PubMed: 20457814]
76. Thomas MC, Chiang C-M. 2006 The general transcription machinery and general cofactors. *Crit. Rev. Biochem. Mol. Biol* 41:105–78 [PubMed: 16858867]
77. Trowitzsch S, Viola C, Scheer E, Conic S, Chavant V, et al. 2015 Cytoplasmic TAF2-TAF8-TAF10 complex provides evidence for nuclear holo-TFIID assembly from preformed submodules. *Nat. Commun* 6:6011 [PubMed: 25586196]
78. Tsai FT, Sigler PB. 2000 Structural basis of preinitiation complex assembly on human pol II promoters. *EMBO J.* 19:25–36 [PubMed: 10619841]
79. Tsai KL, Tomomori-Sato C, Sato S, Conaway RC, Conaway JW, Asturias FJ. 2014 Subunit architecture and functional modular rearrangements of the transcriptional mediator complex. *Cell* 157:1430–44 [PubMed: 24882805]
80. Verrijzer CP, Chen J-L, Yokomori K, Tjian R. 1995 Binding of TAFs to core elements directs promoter selectivity by RNA polymerase II. *Cell* 81:1115–25 [PubMed: 7600579]
81. Wang D, Bushnell DA, Westover KD, Kaplan CD, Kornberg RD. 2006 Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. *Cell* 127:941–54 [PubMed: 17129781]
82. Wang H, Curran EC, Hinds TR, Wang EH, Zheng N. 2014 Crystal structure of a TAF1-TAF7 complex in human transcription factor IID reveals a promoter binding module. *CellRes.* 24:1433–44
83. Wu S-Y, Chiang C-M. 2001 TATA-binding protein-associated factors enhance the recruitment of RNA polymerase II by transcriptional activators. *J. Biol. Chem* 276:34235–43 [PubMed: 11457828]
84. Yakovchuk P, Gilman B, Goodrich JA, Kugel JF. 2010 RNA polymerase II and TAFs undergo a slow isomerization after the polymerase is recruited to promoter-bound TFIID. *J. Mol. Biol* 397:57–68 [PubMed: 20083121]

85. Zhang Z, Boskovic Z, Hussain MM, Hu W, Inouye C, et al. 2015 Chemical perturbation of an intrinsically disordered region of TFIIID distinguishes two modes of transcription initiation. *eLife* 4:e07777

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

direction is marked by an arrow. (*b*) Domain architectures for the subunits of TFIID and TFIIA, and (*c*) domain architectures for TFIIB, TFIIF, TFIIIS, TFIIE, and TFIIH. Solid black outlines surround regions with experimentally determined atomic models of the human protein, whereas dotted lines indicate that only homologous structures exist. Gray backgrounds indicate regions that have been structurally modeled by cryo-electron microscopy within the context of the PIC. The length scale for the schematics in panel *c* is twice of that in panel *b*. Abbreviations: 1iD, TAF1-interacting domain; 2iD, TAF2-interacting domain; 2×Bromo, double bromodomain; 2×Cyclin, double cyclin folds; coiled-C, coiled-coil domain; CTD, C-terminal domain; DRD, DNA-damage recognition domain; Fe-S, iron-sulfur cluster domain; HFD, histone fold domain; HTH, helix-turn-helix; LisH, Lis homology domain; NLS, nuclear localization signal; NTD, N-terminal domain; PH, pleckstrin homology domain; PHD, plant homeodomain; TAFH, TAF homology domain; vWF, von Willebrand factor type A domain; WH, winged helix; ZnF, zinc finger; ZnK, zinc knuckle; ZnR, zinc ribbon.

Figure 2.

The conformational landscape of human TFIID. (a) Following 2D classification of electron microscopy (EM) images of free TFIID into classes showing similar views of the complex (*bottom*), the position of lobe A within each class was measured along the long axis of the BC core, and the fraction of particles (*vertical axis*) with a given lobe A position (*horizontal axis*) was plotted as shown (*green curve*), revealing a bimodal distribution of lobe A positioning (18). Positions to the left of center represent the canonical conformation of TFIID, in which lobe A is near lobe C, whereas those to the right represent the rearranged state of TFIID, in which lobe A is near lobe B. The same analysis was done for TFIID in the presence of TFIIA and promoter DNA (*red dashed curve*), revealing a shift in the conformational equilibrium toward the rearranged state. (b) Cryo-EM 3D reconstructions of TFIID in the canonical (*left*, EMD 2287) and rearranged (*right*, EMD 2282) states were generated using a multi-model refinement strategy, with *ab initio* orthogonal tilt reconstructions as initial references (18). The density corresponding to the BC core (*blue*) stays relatively consistent between the two states, whereas the density for lobe A (*yellow*) is dramatically different. An additional density corresponding to the promoter DNA was observed only in the rearranged state reconstruction, indicating that DNA binds rearranged TFIID. A part of lobe A (lobe A1, *orange*) was later found to contain TFIIA and TATA binding protein. Figure modified from Reference 60 with permission.

Figure 3.

Cryo-electron microscopy (EM) analysis of promoter-bound TFIID. (a) Purification of promoter-bound TFIID-TFIIA complexes (52). Human TFIID complexes, previously affinity-purified from HeLa nuclear extract, are mixed with biotinylated (Bio) promoter DNA and recombinant human TFIIA (orange), and the DNA is then immobilized on magnetic streptavidin (SA) coated beads. Unbound DNA and proteins are washed away, and the purified DNA-bound complexes are then released by a restriction enzyme (RE) that cleaves the DNA at a restriction site placed between the promoter sequence and the biotinylated end. (b) 3 D classification of cryo-EM images of the TFIID-TFIIA-DNA complex reveals the flexibility of TFIID's lobe A2 with respect to the more stable DNA-bound core composed of lobes A1, B, and C. The transparent isosurface is displayed at a lower threshold to enable visualization of weaker densities. (c) A strategy for local classification and refinement within masks around the DNA-bound core (*dashed green line*) or lobe C only (*dashed red line*) resulted in improved resolutions in these regions compared to the unmasked reconstruction (*top*) because of the exclusion of flexible parts of the complex. (*Top*, EMD 3304; *middle*, EMD 3305; *bottom*, EMD 3306.) Figure modified from References 52 and 60 with permission.

Figure 4.

Details of promoter binding by TFIID. (*a*) Current model of promoter-bound TFIID-TFIIA (52). The TATA binding protein (TBP)-TFIIA module engages the upstream promoter (*left*), whereas TAF1 and TAF2 engage the downstream region of the promoter (*middle and right*), including the transcription start site (TSS) (+1). A homodimer of the TAF6 HEAT repeat domain bridges lobes C and B, whereas a helical segment of TAF8 bridges TAF2 with one copy of TAF6. The schematic on the bottom right depicts the super core promoter DNA and illustrates which parts of the promoter are contacted by subunits of TFIID. (*b*) The main promoter-interacting regions of TAF1 (*left*) and TAF2 (*right*), colored by amino acid conservation. Highly conserved positively charged residues within the TAF1 winged helix (WH) that appear to interact with the DNA are depicted as ball-and-stick models. (*c, left*) Surface electrostatics of the TAF1-TAF2 subcomplex, showing that a strongly positively charged patch demarcates the DNA-interacting surface (*blue* = positive, *red* = negative). On the right, a side view of the TAF1-TAF2 complex highlights the network of interactions between TAF1, TAF2, and the downstream DNA. (*d*) Current model for promoter binding by TFIID. On the left, TFIID is in the autoinhibitory canonical state, in which TBP is blocked from binding DNA by the TAND of TAF1. Interactions with promoter DNA and TFIIA (noted with the brackets and asterisk) repress the inhibitory effect of TAF1 and drive TFIID into the rearranged state (*middle*), bringing TBP toward lobe B of TFIID. Interactions with

the promoter are probably initiated by TAF1-TAF2 in the downstream promoter region, placing the upstream promoter DNA in position to be engaged by TBP (*right*).

Abbreviations: DPE, downstream promoter element; MTE, motif ten element; TAF, TBP-associated factor. Panel *a* is adapted from Reference 60, and panels *b-c* are adapted from Reference 52.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 5.

Cryo-electron microscopy (cryo-EM) structures to date of core eukaryotic preinitiation complexes (PICs) in different states during transcription initiation. (a) Cryo-EM reconstructions of human PICs from the Nogales group, and of *Saccharomyces cerevisiae* PICs from (b) the Cramer and (c) the Kornberg groups. The cryo-EM density maps are grouped by states: closed complex (CC), open complex (OC), and initially transcribing complex (ITC). Human reconstructions for CC states include those of TATA binding protein (TBP)-TFIIA-TFIIB-DNA-Pol II (i, EMD 2304), plus TFIIF (ii, EMD 2305), and plus TFIIE (iii, EMD 2306) from Reference 36, as well as core CC (missing TFIIH) (iv, EMD 8135) and full CC (v, EMD 3307) from Reference 37. The rest of the human structures correspond to the core OC (vi, EMD 8136), full OC (vii, EMD 8132), core ITC (viii, EMD 8137), full ITC (ix, EMD 8133), core ITC (without TFIIIS) (x, EMD 8138), and full ITC (with TFIIIS) (xi, EMD 8134) from Reference 37. Reconstructions from the Cramer group correspond to core CC (i, EMD 3383), core OC (ii, EMD 3378) from Reference 63, and minimal ITC that included TBP-TFIIB-Pol II-TFIIE (iii, EMD 2785) from Reference 64. Reconstructions from the Kornberg group are labeled for full CC (i, EMD 2394) from Reference 57 (where the density interpretation in terms of subunits was inconsistent with all other structures published), core CC (ii, EMD 3115), and full CC (iii, EMD 3114) from Reference 58. Color scheme for the core promoter elements within the human DNA construct and protein components of the PIC is shown at the bottom.

Figure 6.

Architecture of the human preinitiation complex. Near-atomic model of the core, TATA binding protein (TBP)-based PIC in the closed complex state shown in two perpendicular views (PDB 5IYA). Key general transcription factor domains are labeled. Abbreviations: Pol II, RNA polymerase II; WH, winged helix; ZR, zinc ribbon.

Figure 7.

Comparison of human and yeast preinitiation complex (PIC) models. The structures are aligned using the rigid part of RNA polymerase II (Pol II) (excluding the clamp and the stalk domains) in both (a) full closed complex (CC) [human, PDB 5IY6 (37); yeast, PDB 5FMF (58)] and (b) core open complex (OC) states [human, PDB 5IYB (37); yeast, PDB 5FYW (63)]. The human models are colored following the same scheme as in the rest of the figures. The yeast models are colored in dark green. In the CC, the yeast TATA binding protein (TBP)-TFIIA module, the Pol II clamp domain, TFIIH as exemplified by XPB, and the path of the downstream DNA are in different relative positions. In the OC (TFIIH not shown, as it was not present in the yeast structure for this stage of initiation), the same differences persist for the TBP-TFIIA module and the DNA path between the yeast and human structures. The Pol II clamp domain is in the same closed position in both structures.

Figure 8.

Comparison of human and yeast general transcription factor structure models. (a) Alignment of the TATA binding protein (TBP) TFIIA modules of the human and yeast (dark green) open complex (OC) states shows that the DNA path and the Rap30 winged helix (WH) domain are practically superimposable, except for the presence of an additional short helix at the C-terminal end of Tfg2 [human, PDB 5IYB (37); yeast, PDB 5FYW (63)]. (b) When both subunits of TFIIF are aligned for the human and yeast OC states, the dimerization domain and the overall path of the linker within Rap30 (Tfg2 in yeast) are in similar positions. In contrast, the C-terminal end of the linker and the WH domain of Rap30 deviate significantly between the two structures, as does the DNA path [human, PDB 5IYB (37); yeast, PDB 5FYW (63)]. (c) When both subunits of TFIIE are aligned between human and yeast CC states, the TFIIEα helix-turn-helix and WH are superimposable. The clamp coiled-coil domain of Pol II interacts with TFIIE using different interfaces for the yeast and human systems. The beta hairpin in the TFIIEα WH domain interacts with the duplex DNA in the yeast structure, whereas it is partially disordered (and thus lacking interaction with the DNA) in the human structure [human, PDB 5IYA (37); yeast, PDB 5FZ5 (63)]. (d) Alignment of TFIIB between the human and yeast OC states shows that the two cyclin folds are superimposable, but the TFIIB linker region is stabilized only in the human structure, where it directly interacts with the nontemplate DNA strand [human, PDB 5IYB (37); yeast, PDB 5FYW (63)]. (e) Modeled regions of TFIIF are aligned between the human and yeast CC states. TBP is in a very similar position, indicating that the distance between the XPB-

DNA interacting site and the TATA box is the same between the yeast and human structures [human, PDB 5IY6 (37); yeast, PDB 5FMF (58)].

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 9.

Structural transitions during promoter opening. (a) Human closed complex (CC) and open complex (OC) structures are aligned by superposition of the rigid part of RNA polymerase II. The DNA near the XPB binding site is in the same position. However, (b) the fork loop 2 and (c) the clamp-TFIIE regions undergo significant changes. (d) TFIIE becomes closer to the rest of the preinitiation complex in the OC, as it pivots around the point of XPB-DNA contact [CC, PDB 5IY6; OC, PDB 5IY7 (37)]. The OC model is colored following the same scheme as in the rest of the figures, and the CC model is colored in dark green.

Figure 10.

Model of the human TFIID-based preinitiation complex (PIC). (a) Superposition of the common elements in the human TATA-binding protein (TBP)-based PIC (*upper left*, EMD 3307) and human TFIID-TFIIA-DNA complex (*lower left*, EMD 3305), including TBP, TFIIA, and the promoter DNA (*ribbon representation*), results in a synthetic structural model for the human TFIID-based PIC (52). The model suggests that minor rearrangements are required within the TFIID-TFIIA-DNA complex to accommodate RNA polymerase II (Pol II) into the PIC and supports previous findings that subunits of TFIID can interact with other general transcription factors (GTFs), including TFIIF and TFIIE. On the bottom right, a

depiction of the super core promoter DNA illustrates which parts of the promoter are contacted by GTFs. (b) Putative mechanism for Pol II recruitment to the TFIID-TFIIA-promoter complex. Following the loading of TBP onto the upstream promoter DNA (see Figure 4d), interactions between Pol II-bound TFIIB and the TBP-DNA complex likely mediate the recruitment of Pol II to the complex. Following the engagement of promoter DNA by TFIIF and Pol II (*middle*), at least some of the downstream promoter contacts with TFIID are released. TFIIE and TFIIH are then recruited to the PIC (*right*). It is yet unclear whether TAFs remain bound to the PIC in the final stages of PIC assembly or through transcription initiation. (c) Superposition of current cryo-electron microscopy structures from both human and yeast systems yields a model of an ~3 MDa transcription initiation supracomplex including TFIID, TFIIA, TFIIB, TFIIF, TFIIE, TFIIH, Pol II, and the Mediator coactivator complex (79). The flexible lobe A of TFIID (*yellow*) is depicted at two different isosurface thresholds, with lower threshold in transparency approximating its range of positioning. Figure modified from Reference 52 with permission.