

Structure of an RNA polymerase II preinitiation complex

Kenji Murakami^{a,b,1}, Kuang-Lei Tsai^{c,1}, Nir Kalisman^d, David A. Bushnell^a, Francisco J. Asturias^c, and Roger D. Kornberg^{a,2}

^aDepartment of Structural Biology, Stanford University School of Medicine, Stanford, CA 94305; ^bDepartment of Biochemistry and Biophysics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104; ^cDepartment of Integrative Structural & Computational Biology, The Scripps Research Institute, La Jolla, CA 92037; and ^dDepartment of Biological Chemistry, Institute of Life Sciences, The Hebrew University of Jerusalem, Jerusalem 91904, Israel

Contributed by Roger D. Kornberg, September 22, 2015 (sent for review August 31, 2015; reviewed by Arnold J. Berk and Jesper Q. Svejstrup)

The structure of a 33-protein, 1.5-MDa RNA polymerase II preinitiation complex (PIC) was determined by cryo-EM and image processing at a resolution of 6–11 Å. Atomic structures of over 50% of the mass were fitted into the electron density map in a manner consistent with protein–protein cross-links previously identified by mass spectrometry. The resulting model of the PIC confirmed the main conclusions from previous cryo-EM at lower resolution, including the association of promoter DNA only with general transcription factors and not with the polymerase. Electron density due to DNA was identifiable by the grooves of the double helix and exhibited sharp bends at points downstream of the TATA box, with an important consequence: The DNA at the downstream end coincides with the DNA in a transcribing polymerase. The structure of the PIC is therefore conducive to promoter melting, start-site scanning, and the initiation of transcription.

transcription | general transcription factors | yeast | cryo-EM

Five general transcription factors (GTFs), termed TFIIB, -D, -E, -F, and -H, are required for the initiation of RNA polymerase II (pol II) transcription. The GTFs can be assembled with pol II and promoter DNA in a preinitiation complex (PIC) capable of efficient conversion to a transcribing complex (0.1–0.3 transcripts per template). Cryo-EM of the PIC at a resolution of 15–20 Å revealed a bipartite structure, with a P-lobe containing pol II and a G-lobe containing most of the mass of the GTFs (1). A cylinder of electron density, attributed to promoter DNA, was in contact only with GTFs and not with pol II. This architecture was consistent with the pathway of assembly of the PIC, beginning with a complex of promoter DNA and all but one of the GTFs, to which a complex of pol II and the remaining GTF was added (2).

A combination of chemical cross-linking and mass spectrometry was used to locate the subunits of the GTFs in the low-resolution cryo-EM electron density map of the PIC. TFIIB was in proximity to promoter DNA at the upstream end of the pol II active center cleft, whereas Ssl2, a subunit of TFIIF, was in apparent contact with promoter DNA at the downstream end of the cleft. TFIIB bridges between pol II and promoter DNA, whereas Ssl2 is a DNA helicase, responsible for unwinding promoter DNA.

The association of promoter DNA with the GTFs and not with pol II may be viewed as a fundamental principle of the PIC. Double-stranded DNA is straight and relatively rigid, whereas DNA would need to bend by about 90° to bind in the pol II active center cleft. The GTFs assemble on the double-stranded DNA, position it above the active center, and unwind the DNA. The resulting single-stranded region is flexible and can bend and bind in the pol II cleft.

Cryo-EM of the PIC with state-of-the-art instrumentation has resulted in an electron density map at higher resolution. The improved map has confirmed and extended the previous findings and conclusions.

Results

Cryo-EM and 3D Reconstruction. A PIC was formed as described (2) on an 86-bp fragment of *HIS4* promoter DNA fragment, with pol II,

GTFs (TFIIA, TFIIB, TBP, TFIIE, TFIIIF, and holo-TFIIF), TFIIS, and with the addition of Sub1 (yeast homolog of PC4, which stimulates the initiation of transcription) (3, 4). The PIC was sedimented in a glycerol gradient in the presence of a non-hydrolyzable analog of ATP. Peak gradient fractions contained equimolar amounts of all transcription factors and a dimer of Sub1 (5). Cryo-EM was performed with a Titan Krios (FEI), equipped with a K2 Summit direct electron detector (Gatan). Approximately 74,000 images of PIC particles were aligned and clustered with the SPARX EM image processing package (6) and the resulting class averages were used for ab initio calculation of an initial PIC map at about 25-Å resolution (Fig. S1). The map showed a clear division in two parts, termed G-lobe and P-lobe (Fig. 1), as previously observed (1). A crystallographic model of a pol II–TFIIB–TBP–TATA DNA fragment (7, 8) with TFIIA (9) was docked to the P-lobe. The B-finger and B-linker of TFIIB (8, 10) were not resolved in the map. There was a close correspondence of secondary structural features of pol II, of the shapes of the GTFs, and of the DNA double helix to the map. The close fit of the crystallographic model to the P-lobe portion validated the initial map as a starting point for more detailed analysis of the PIC structure.

Focused image classification with SPARX revealed three forms of the PIC (Fig. S2). Forms 1 and 2, accounting for about 20% and 10% of the images, respectively, showed well-ordered P- and G-lobes, but differed in the position of the G-lobe. In the remaining 70% of the images, the G-lobe was variable in structure or location. Further image alignment and classification with the RELION EM image processing package (11) identified about 7,000 images that

Significance

RNA polymerase II, the enzyme responsible for all mRNA synthesis in eukaryotes, requires a set of general transcription factors (GTFs) for the initiation of transcription. A complex of the polymerase and GTFs, with a mass of 1.5 MDa, was previously isolated and shown to be stable, homogeneous, and active in transcription. A cryo-electron microscope structure of the complex at a resolution of 6–11 Å reported here explains the requirement for the GTFs and elucidates their roles. The GTFs recognize the region of the gene responsible for initiation and deliver it to the polymerase active center in a form amenable to transcription.

Author contributions: K.M., K.-L.T., N.K., F.J.A., and R.D.K. designed research; K.M., K.-L.T., N.K., and F.J.A. performed research; K.M., K.-L.T., N.K., D.A.B., F.J.A., and R.D.K. analyzed data; and K.M., K.-L.T., and R.D.K. wrote the paper.

Reviewers: A.J.B., Molecular Biology Institute; and J.Q.S., Cancer UK.

The authors declare no conflict of interest.

Data deposition: The data reported in this paper have been deposited in the Electron Microscopy Data Bank [accession nos. EMD-3114 (preinitiation complex) and EMD-3115 (refined P-lobe)].

¹K.M. and K.-L.T. contributed equally to this work.

²To whom correspondence should be addressed. Email: kornberg@stanford.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1518255112/-DCSupplemental.

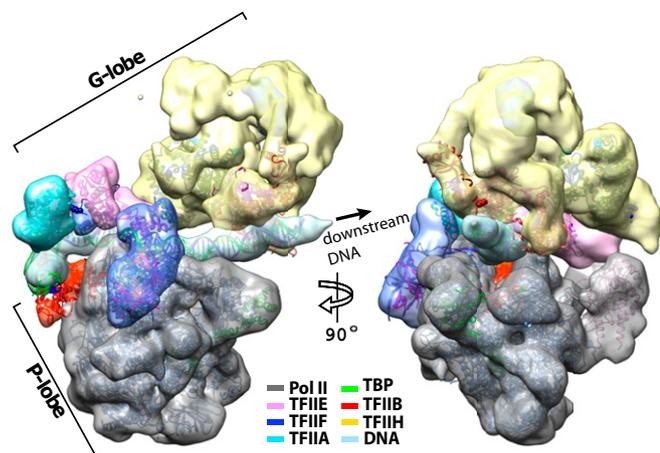


Fig. 1. Cryo-EM structure of the PIC. Side view (Left) and front view (Right) of the cryo-EM electron density map (gray, space-filling) are shown. A crystallographic model of a pol II (all light blue, except Rpb4/7 subunits pink)–TBP (green)–TFIIB_C (red)–TATA box complex and crystal structures of TFIIA (cyan) and TFIIS (green) were fitted to the map. Atomic models of the Tfg1–Tfg2 dimerization domain (Tfg1 blue; Tfg2 magenta), the WH domain of Tfg2 (navy blue), WH domains of Tfa2 (magenta), the WH domain of Tfa1 (purple), the zinc ribbon domain of Tfa1 (navy blue), Rad3 (green), Ssl2 (N-Ssl2 orange red; C-Ssl2 dark purple), and Tfb2–Tfb5 (dark red) were placed on the basis of protein–protein cross-linking. Volumes assigned on the basis of fitting the atomic models and results of cross-linking are colored according to the code at the lower left.

were used to refine a map of form 1 to an overall resolution of about 11 Å (Fig. 1). This map is described in detail below, and differences from a map of form 2, calculated from about 4,000 images, are discussed as well. The structure of the P-lobe was more consistent throughout the dataset than that of the G-lobe, and a map of the P-lobe at a resolution of about 6 Å was obtained by masking the G-lobe and refinement from subset of about 37,000 images. The only significant difference between the EM map and the crystallographic model of a pol II–TFIIB–TBP–TATA DNA complex (7, 8) with TFIIA (9) was a rotation of about 20° of the pol II subunits Rpb4 and Rpb7 (Fig. S3).

Model Building of the PIC. Placing proteins in the G-lobe region of the EM map of form 1 was guided by previous results of chemical cross-linking and mass spectrometry (1, 12). Cross-links were related to the electron density as described (1). Briefly, 12 peaks of highest electron density at least 45 Å apart in the map were identified in the G-lobe. The two subunits of TFIIE (Tfa1 and Tfa2) and six subunits of TFIH (Tfb1, Tfb2, Tfb4, Tfb5, Ssl1, and Ssl2) were divided into 12 globular domains, based on the results of cross-linking and homology to crystal structures. All possible placements of the globular domains at the 12 peaks of electron density in the map were evaluated with respect to the pattern of cross-links, and a single best-fitting model was obtained (Fig. S4). This coarse-grained model is very similar in the arrangement of subunits to that in our previous analysis (1), differing primarily by a translation of about 35 Å in the locations of TFIH subunits, due to a shift in the location of electron density due to TFIH (Fig. S5).

A crystallographic model of DNA bound to XPB, a human homolog of Ssl2 (13), was a good match to EM density (Fig. 2*A* and *B*). The best fit was found by moving the XPB model along the DNA, retaining the association with the minor groove, characteristic of this family of helicases (14). The crystal structure of human XPD, a human homolog of Rad3 (15), was a good match to EM density in the location identified by cross-linking for Rad3 (Fig. 2*C* and *D*). Manual fitting of the crystal structure was optimized by rigid-body refinement. The resulting model explains

why residues K95 and K731 of Rad3 (colored cyan in Fig. 2*C* and *D*) both form cross-links to the core subunits of TFIH (1, 16). Although Rad3 was in close proximity to TFIIE, there was little or no contact between them (Fig. 2*C* and *D*). EM density bridging between Rad3 and Tfa1 may be due to Tfb3, which interacts with the ARCH domain of Rad3 (17); EM density adjacent to the ARCH domain is indicated by a dashed circle in Fig. 2*D*. The crystal structure of Tfb2–Tfb5 (18) was docked into a corresponding density adjacent to the C-terminal half of Ssl2. The fit explained why K721 of Ssl2 could be cross-linked to K60 of Tfb5 (Fig. 2*B*) (1, 16).

Path of DNA in the PIC. Rod-like EM density was identifiable as DNA on the basis of both the diameter and apparent grooves of the double helix (Fig. 2*A* and *B*). The density extends from about 15 bp upstream of the bend at the TATA box to about 55 bp of DNA downstream (from –75 to –5 relative to the transcription start site at +1). There are four notable features of the path of the DNA (Figs. 2 and 3). First, there is no contact of the DNA with pol II (see especially Fig. 3*A*, front view). Second, the DNA is suspended above the pol II cleft, bound by TBP, TFIIB, TFIIA, and Tfg2 at the upstream end and by Ssl2 at the downstream end; it is free in between. Third, the DNA downstream of the TATA box is bent, best modeled by three segments of B-form double helix (Fig. 3*A*). Finally, the segment of double helix bound to Ssl2, about 45 bp downstream of the TATA box, is in the position and orientation of DNA entering a pol II transcribing complex (Fig. 3*B*, showing straight B-form DNA added at the downstream end of the transcribing complex).

Refinement of the P-Lobe. The 6-Å map of the P-lobe contained pol II, TFIIS, TFIIB, TBP, TFIIA, and part of TFIIE. About 20 bp of DNA upstream of the TATA box and about 30 bp downstream are also present (Fig. 4). DNA density becomes partially disordered after the first bend beyond the TATA box in the DNA of the complete PIC. Structures of the Tfg1–Tfg2 dimerization domain and the Tfg2 winged helix (WH) domain (19, 20) could be fit to the EM density and their positions optimized by rigid body refinement (Fig. 4). The WH domains of TFIIE [Tfa2 WH1 (residues 125–184), WH2 (residues 185–247), and Tfa1 WH (residues 12–95)] were placed in the EM density on the basis of cross-linking and protein–DNA mapping studies (dashed lines in Fig. 4*B*). The fitting of the two WH domains of Tfa2 was also constrained by the short linker between them. Density adjacent to the Tfa1–Tfa2 dimerization domain but not accounted for by the WH domains was attributed to the C-terminal helix of Tfa2 (residues 251–267, labeled C-Tfa2 in Fig. 4*B*), near the DNA at about position –40, and to the zinc ribbon domain of Tfa1, near Rpb4/7.

Discussion

The model from cryo-EM of the PIC at a resolution of 6–11 Å confirms the main conclusions of the previous study at a resolution of 15–20 Å (1): The PIC is divided into a P-lobe and G-lobe; the promoter DNA is not in contact with pol II, but rather is associated with GTFs, including the helicase for unwinding the DNA; and the GTFs thereby deliver promoter DNA to pol II in an unwound form suitable for binding in the active center and transcription. The model extends the previous results by the placement of near-atomic structures of more than 50% of the protein mass of the PIC and by definitive localization of the promoter DNA. The model satisfied 87 of 90 high-significance cross-links (C α –C α distances less than 40 Å, the span of the cross-linking reagent). Two of the remaining three cross-links, between Rpb6 and the Tfg1–Tfg2 dimerization domain, were only slightly aberrant (C α –C α distances on average 49 Å) and were likely due to flexibility of the regions involved. The remaining cross-link, at a C α –C α distance of 60 Å, was likely

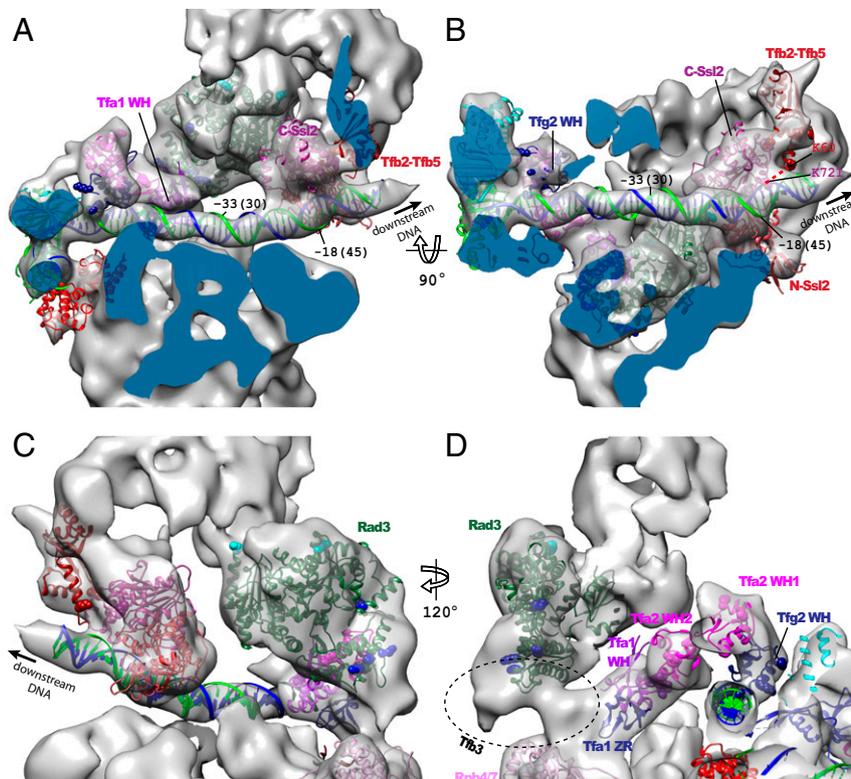


Fig. 2. Placement of crystal structures for TFIIH subunits. (A and B) Density has been removed (cut surfaces dark blue) to expose N- and C-terminal regions of Ssl2 (N-Ssl2 orange red; C-Ssl2 dark purple). Also shown are Tfb2–Tfb5 (dark red) and the WH domains of Tfa1 (pink) and Tfg2 (navy blue). (C and D) Rad3 (green). Volumes are colored as in Fig. 1. Residues of Rad3 in cyan form cross-links with core subunits of TFIIH, and residues in blue form cross-links with Tfa2 in PIC lacking TFIK. If the cross-linked residue is absent from the model, the closest residue in the model is shown. Cross-link between K60 of Tfb2 and K725 of Ssl2 is indicated by a dashed red line. Positions in the DNA are numbered with respect to the first transcription start site of the *HIS4* promoter, with distances (base pairs) from the upstream edge of the TATA box in parentheses.

misassigned (consistent with the 1.5% false-positive rate of the cross-linking analysis).

Nine of the 33 proteins of the PIC are not modeled or assigned. Four core subunits of TFIIH in the G-lobe (Tfb1, the N-terminal part of Tfb2, Tfb4, and Ssl1), which amount to about

10% of the total mass of the PIC, are not modeled into corresponding EM densities. Five subunits (Ccl1, Kin28, Tfb3 of TFIK, Tfg3 of TFIIF, and Sub1), which amount to another about 10% of the total mass of the PIC, lacked corresponding densities. A density between Rad3 and Tfa1 may be attributed to

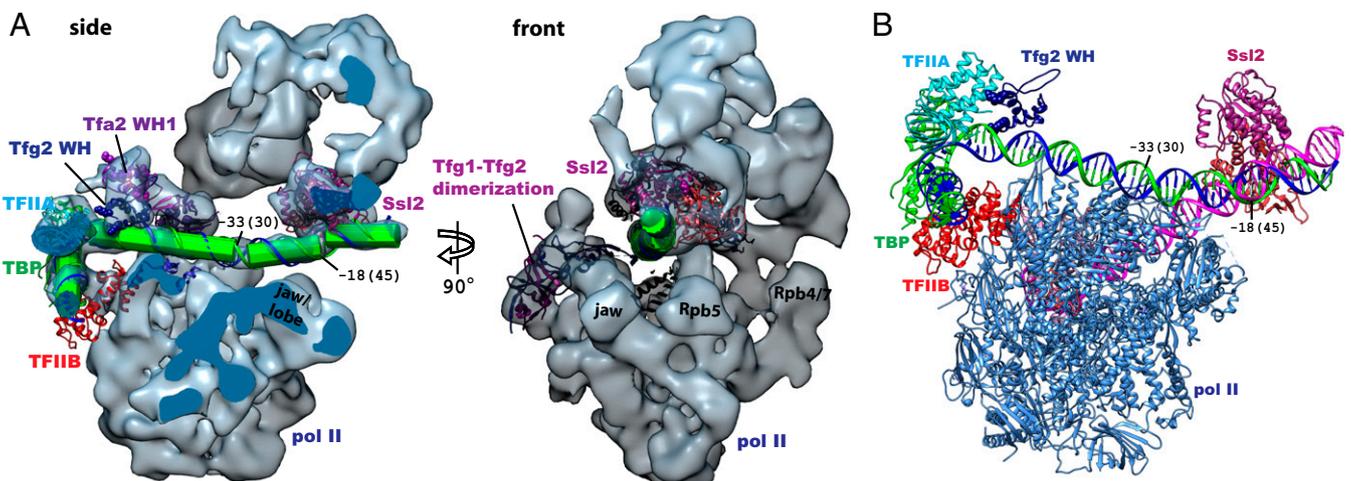


Fig. 3. Path of promoter DNA in the PIC. (A) Views of DNA path, with green cylinders representing straight B-form DNA segments superimposed. Proteins contacting the DNA are shown: TFIIIA (cyan), TBP (green), TFIIIB (red), Ssl2 (N-Ssl2 orange red; C-Ssl2 dark purple), WH domains of Tfa2 (magenta), Tfg2 (navy blue), and Tfa1 (purple), and the zinc ribbon domain of Tfa1 (navy blue). Positions in the DNA are numbered as in Fig. 2. (B) Pol II in the crystal structure of a transcribing complex (40) is aligned with pol II in the PIC, the DNA of the transcribing complex (magenta) is extended at the downstream end with straight B-form DNA, and the DNA of the PIC with all associated proteins is shown as in Fig. 1.

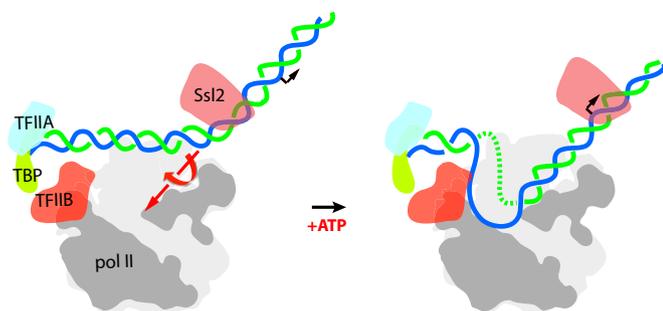


Fig. 5. Schematic of transition from closed to open promoter complex. Cutaway views based on Fig. 3B. TFIIE, TFIIF, and the core subunits of TFIIF are omitted for clarity. DNA in blue and green is rotated and translocated by Ssl2 in an ATP-dependent manner, in the directions indicated by the large red arrows.

Finally, it may be noted that the distribution of electron density in form 1 resembles that of He et al. (27) obtained by negative staining of a human complex. Although there is little difference in the general picture between our work and theirs, differences in the DNA path, the locations of GTFs, and interpretation remain to be resolved.

Materials and Methods

Protein Purification and PIC Preparation. TFIIA, TFIIB, TBP, TFIIS, and Sub1 were available in recombinant form, and TFIIE, TFIIF, TFIIF- Δ TFIIF, TFIIF, and pol II were isolated from yeast as previously published (26). The PIC was prepared as previously described (2) with minor modifications; 0.6 nmol of a *HIS4* promoter DNA fragment ($-85/+1$) was mixed with 1.0 nmol of TFIIB, 0.9 nmol of TFIIA, 0.7 nmol of TBP, 1.0 nmol of TFIIE, 0.4 nmol of TFIIF- Δ TFIIF, 0.8 nmol of TFIIF, 0.6 nmol of sub1 in 130 μ L of buffer (500) [20 mM Hepes (pH 7.6), 5 mM DTT, 2 mM Mg(OAc)₂, and 5% glycerol, with the millimolar concentration of KOAc in parentheses]. The mixture was dialyzed into buffer (300), buffer (220), buffer (150), and then combined with 0.33 nmol of pol II, 0.66 nmol of TFIIF, and 1.2 nmol of TFIIS. The mixture was further dialyzed into buffer (90), buffer (40), and was loaded onto a 10–40% (vol/vol) glycerol gradient containing 20 mM Hepes (pH 7.6), 5 mM DTT, 2 mM Mg(OAc)₂, 40 mM KOAc, and 500 μ M AMP-PNP and was centrifuged for 8 h at 36,000 rpm in a Beckman SW60 Ti rotor. For cryo-EM, PICs were fixed by sedimentation in glycerol gradients containing a gradient of glutaraldehyde from 0 to 0.125% (28). Aliquots of peak fractions (\sim 0.3 mg/mL) were flash-frozen in liquid nitrogen and stored until use at -80 °C.

Specimen Preparation and EM. Twenty microliters of glutaraldehyde-fixed sample (\sim 0.3 mg/mL) was dialyzed in a Slide-A-Lyzer MINI Dialysis Unit (20,000 molecular weight cutoff) (Thermo Scientific) against 20 mM Hepes (pH 7.6), 5 mM DTT, 2 mM Mg(OAc)₂, and 40 mM KOAc for 2–3 h to remove glycerol. Electron microscope grids (Quantifoil R2/1) were glow-discharged for 15 s using a Harrick Plasma Cleaner (PDC-32G) before application of \sim 3 μ L PIC sample and then were flash-frozen in liquid ethane using a Vitrobot (FEI) at 100% humidity and 22 °C (blot time 4 s, waiting 10 s, force 9).

Imaging and Preliminary Image Processing. Cryo-EM specimens were imaged using a Titan Krios microscope (FEI) operating at 300 kV, at a magnification of 22,500 \times (resulting in image sampling at 1.315 Å per pixel), and with underfocus values between 1 and 3 μ m. Specimen images were recorded on a K2 Summit direct electron detector (Gatan) operated in counting mode using Legion. Total electron dose for each image was 40 electrons Å^{-2} over a 7-s exposure time, fractionated into 35 frames. Dose-fractionated frames were aligned (29) and contrast transfer function parameters for each image were determined using the program sxcter (SPARX). About 10,000 particle images were manually selected and used to calculate 2D class averages using the ISAC clustering algorithm, which uses alignment parameter stability and clustering reproducibility as criteria to group particle images into homogeneous classes (30). The best 2D class averages from ISAC were then used as templates for automated picking of 113,825 particle images using Appion (31). An initial step of image screening was performed using the clustering protocol implemented within Appion. A second round of screening was performed using ISAC (SPARX) on fourfold decimated images (5.26 Å per pixel). Class averages lacking clear features, or not resembling possible projections of a macromolecular complex,

were eliminated. The remaining averages (including a total of about 74,000 particle images) were used to calculate an initial 3D model of the PIC using the program VIPER (SPARX). The screened images were processed with 3D auto-refinement, movie refinement, and particle polishing routines in RELION, and the resulting “shiny” images were used for all further image processing steps. Initial alignment parameters for all shiny images were determined by deterministic matching to projections of the initial VIPER model using SPARX.

Identification of Optimal Particle Image Subsets. The initial map of the PIC calculated ab initio using VIPER showed a structure consisting of two lobes surrounding a central channel. Most of the mass in one of the lobes (P-lobe) clearly corresponded to pol II, implying that TFIIF would most likely account for most of the mass in the other half (G-lobe). Examination of the ISAC averages after considering the initial PIC model from VIPER indicated that the P-lobe was generally well-ordered, whereas the other half of the PIC varied in conformation. To obtain the best possible 3D maps of the entire PIC and each of its lobes, particle images were clustered using two independent approaches. First, 3D classification using RELION was used to identify various image subsets. Most images were assigned to groups showing a well-resolved P-lobe next to a poorly ordered G-lobe. However, two smaller image subsets corresponded to two well-ordered, but different, conformations of the PIC. A subset of about 7,000 images (7K_PIC stack) resulted in the best 3D map of the entire PIC we could obtain (form 1). A smaller \sim 4,000-image stack (4K_PIC stack) also showed the entire PIC, but with TFIIF in a different conformation (form 2). We also used SPARX to pursue a different approach in which parameters required to center the P- and G-lobe portions of the initial 3D model were determined and transformed as required for centering the P- and G-portions of each particle image. Application of these 2D centering parameters resulted in transformed stacks in which either the P- or G-lobe was centered. Images in these two transformed stacks were again clustered with ISAC to identify about 37,000 images with a centered, well-defined P-lobe (37K_P stack), and about 30,000 images with a centered, well-defined G-lobe (30K_H stack).

Refinement of Image Parameters and Calculation of Final Cryo-EM Maps. SPARX refinement of a twofold decimated (2.63 Å per pixel) 7K_PIC stack using the VIPER 3D map as initial model resulted in a cryo-EM map of the PIC with a resolution of about 11 Å, in which the DNA running down the middle of the PIC appeared nicely resolved. This 7K_PIC SPARX map was used for interpretation of the overall PIC structure.

The 7K_PIC SPARX map was used as template to generate a 3D mask that included density corresponding to the P-lobe (pol II, TFIIA, TFIIB, TBP, TFIIF, TFIIE, promoter DNA, and the portions of TFIIF in direct contact with the P-lobe). This mask was applied during refinement of alignment parameters for the particle images in the 37K_P stack using SPARX. This resulted in a map with optimal definition of features in the P-lobe map portion of the PIC structure. The overall resolution of the P-lobe map was about 6 Å, although peripheral portions of the map (corresponding to GTFs and Rpb4–Rpb7) have lower resolution, most likely resulting from some local mobility (Fig. S1).

A similar procedure was followed to determine a G-lobe map from the images in the 30K_H stack. However, the considerable variability in organization of TFIIF already apparent in ISAC averages of the PIC prevented refinement of a map past about 15-Å resolution. Therefore, the best structural description of the G-lobe we could obtain actually came from the 7K_PIC_SPARX map.

Combinatorial Analysis of Protein–Protein Cross-Links. The approximate positions within the electron density of subunits of TFIIE and TFIIF were first estimated on the basis of cross-linking data as previously described (1). Briefly, 12 positions were marked within the unassigned electron density by an iterative and greedy algorithm. The highest-density voxel was taken as the first location. The second location was chosen as the highest-density voxel that was farther than 45 Å from the first location. The third location was the highest-density voxel that was farther than 45 Å from the first and second locations. This procedure was repeated 12 times. Following this greedy iteration, the spheres were allowed to move up to 10 Å from their initial position to optimize the electron density they occupy. These positions span the density and parse it into subregions of roughly equal volume. In parallel, the sequences of the two subunits of TFIIE and seven core subunits of TFIIF were parsed into 12 domains, which were assumed to be compact and globular, based on cross-linking data or homology to crystal structures (see figure S12 in ref. 1). There are 12 factorial (480 million) different models that assign the 12 spheres to the 12 positions, and we assessed them exhaustively. First, we discarded models in which two spheres belonging to the same

protein were more than 65 Å apart, reducing the number of models to a million. We then evaluated the fit of each model to the pattern of cross-links on the basis of two measures: serious violations, defined as the number of pairs of spheres located more than 65 Å apart in the model, for which cross-links are nevertheless observed, and violation distance, defined as the total excess over 65 Å of pairs of spheres in the model for which cross-links are observed. These two measures were correlated over a wide range of values (Fig. S4B). The threshold of 65 Å represents the maximal bridging distance of the BS3 cross-linking reagent (30 Å) plus an assumed radius for each of the 12 domains (35 Å), and the results did not vary for similar thresholds that were tried. If a pair of spheres was connected by several cross-links, only one cross-link instance was considered to avoid bias toward specific pairs of spheres. The model with the smallest sum of violation distances (Fig. S4A) is similar to the one published (1) with regard to the positions of the Tfa1, Tfa2, Rad3, Tfb2/Tfb5, and Ssl2 subunits. These subunits are also most frequently assigned to the same positions in the 20 best-fitting models (Dataset S1), a consensus that is indicative of the confidence of subunit assignment.

Homology Modeling. Yeast homology models of TFIIe, TFIIIF, and TFIIH were generated as follows. A yeast homology model of the Tfg1–Tfg2 dimerization domain was generated using the crystal structure of human TFIIIF (19). A yeast homology model of the Tfg2 WH domain was generated using the NMR structure of the small subunit of human TFIIIF (32). A yeast homology model of the Tfa2 WH domain 1 was generated using the NMR structure of human TFIIeβ (33). A yeast homology model of the Tfa2 WH domain 2 was generated using PDB ID code 1BM9 as template. A homology model of the WH domain of Tfa1 was generated using the N-terminal domain of TFE, the

archaeal homolog of Tfa1 (34). A yeast homology model of the Tfa1 zinc ribbon domain was generated using the NMR structure of human TFIIEx (35). A yeast homology model of Rad3 was generated using the archaeal XPD (15). A yeast homology model of DNA-Ssl2 was obtained using a crystallographic model of DNA-XPB (a human homolog of Ssl2) (13).

Model Building of the PIC. A crystallographic model of a pol II-TBP-TFIIIB-DNA complex, derived from X-ray structures of pol II-TFIIIB and of C-terminal fragments of TBP and TFIIIB bound to a TATA box DNA fragment (7, 8, 36), was fitted into the EM map (37) without any deviations except a 20° rotation of Rpb4/7 subunits. The promoter DNA was extended with B-form DNA, manually fitted into the EM map, and refined using Coot (38). Crystal structures were fitted manually and then computationally as rigid bodies into the EM map (37). The final models of Rad3, Ssl2, and the Tfg1–Tfg2 dimerization domain were chosen based on the fitting score (number of atoms within the EM map) (37). For the modeling of Tfb2–Tfb5 (18), the Tfg2 WH domain, Tfa2 WH domains, and the Tfa1 zinc ribbon domain, the computational fitting gave us two or three top candidates, and the final model was chosen based on XL-MS (1) [or Fe-BABE (39) for the Tfa1 WH domain].

ACKNOWLEDGMENTS. We thank Dr. Katsuhiko Murakami (Pennsylvania State University) for help with modeling of TFIIe, Dr. Jeffrey Ranish (Institute for Systems Biology) for help and discussions of XL-MS, Dr. Holger Stark for help with specimen preparation for cryo-EM, and Dr. E. Peter Geiduschek for critical reading of the manuscript. This research was supported by NIH Grants AI21144 and GM49985 (to R.D.K.) and GM067167 (to F.J.A.). Yeast fermentation was performed with an instrument funded by NIH S10 shared instrumentation Grant S10RR028096 (to R.D.K.).

- Murakami K, et al. (2013) Architecture of an RNA polymerase II transcription pre-initiation complex. *Science* 342(6159):1238724.
- Murakami K, et al. (2013) Formation and fate of a complete 31-protein RNA polymerase II transcription preinitiation complex. *J Biol Chem* 288(9):6325–6332.
- Ge H, Roeder RG (1994) Purification, cloning, and characterization of a human coactivator, PC4, that mediates transcriptional activation of class II genes. *Cell* 78(3):513–523.
- Henry NL, Bushnell DA, Kornberg RD (1996) A yeast transcriptional stimulatory protein similar to human PC4. *J Biol Chem* 271(36):21842–21847.
- Branden J, et al. (1997) C-terminal domain of transcription cofactor PC4 reveals dimeric ssDNA binding site. *Nat Struct Biol* 4(11):900–903.
- Hohn M, et al. (2007) SPARX, a new environment for Cryo-EM image processing. *J Struct Biol* 157(1):47–55.
- Kostrewa D, et al. (2009) RNA polymerase II-TFIIIB structure and mechanism of transcription initiation. *Nature* 462(7271):323–330.
- Liu X, Bushnell DA, Wang D, Calero G, Kornberg RD (2010) Structure of an RNA polymerase II-TFIIIB complex and the transcription initiation mechanism. *Science* 327(5962):206–209.
- Tan S, Hunziker Y, Sargent DF, Richmond TJ (1996) Crystal structure of a yeast TFIIIA/TBP/DNA complex. *Nature* 381(6578):127–151.
- Bushnell DA, Westover KD, Davis RE, Kornberg RD (2004) Structural basis of transcription: An RNA polymerase II-TFIIIB cocrystal at 4.5 Å resolution. *Science* 303(5660):983–988.
- Scheres SH (2012) RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* 180(3):519–530.
- Kalishman N, Adams CM, Levitt M (2012) Subunit order of eukaryotic Tric/CCT chaperonin by cross-linking, mass spectrometry, and combinatorial homology modeling. *Proc Natl Acad Sci USA* 109(8):2884–2889.
- Hilario E, Li Y, Nobumori Y, Liu X, Fan L (2013) Structure of the C-terminal half of human XPB helicase and the impact of the disease-causing mutation XP11BE. *Acta Crystallogr D Biol Crystallogr* 69(Pt 2):237–246.
- Dürr H, Flaus A, Owen-Hughes T, Hopfner KP (2006) Snf2 family ATPases and DExx box helicases: Differences and unifying concepts from high-resolution crystal structures. *Nucleic Acids Res* 34(15):4160–4167.
- Fan L, et al. (2008) XPD helicase structures and activities: Insights into the cancer and aging phenotypes from XPD mutations. *Cell* 133(5):789–800.
- Luo J, et al. (2015) Architecture of the human and yeast general transcription and DNA repair factor TFIIH. *Mol Cell* 59(5):794–806.
- Abdulrahman W, et al. (2013) ARCH domain of XPD, an anchoring platform for CAK that conditions TFIIH DNA repair and transcription activities. *Proc Natl Acad Sci USA* 110(8):E633–E642.
- Kainov DE, Vitorino M, Cavarelli J, Poterszman A, Egly JM (2008) Structural basis for group A trichothiodystrophy. *Nat Struct Mol Biol* 15(9):980–984.
- Gaiser F, Tan S, Richmond TJ (2000) Novel dimerization fold of RAP30/RAP74 in human TFIIIF at 1.7 Å resolution. *J Mol Biol* 302(5):1119–1127.
- Kilpatrick AM, Koharudin LM, Calero GA, Gronenborn AM (2012) Structural and binding studies of the C-terminal domains of yeast TFIIIF subunits Tfg1 and Tfg2. *Proteins* 80(2):519–529.
- Kim TK, Ebright RH, Reinberg D (2000) Mechanism of ATP-dependent promoter melting by transcription factor IIH. *Science* 288(5470):1418–1422.
- Fishburn J, Tomko E, Galburt E, Hahn S (2015) Double-stranded DNA translocase activity of transcription factor TFIIH and the mechanism of RNA polymerase II open complex formation. *Proc Natl Acad Sci USA* 112(13):3961–3966.
- Dvir A, Conaway RC, Conaway JW (1997) A role for TFIIH in controlling the activity of early RNA polymerase II elongation complexes. *Proc Natl Acad Sci USA* 94(17):9006–9010.
- Spangler L, Wang X, Conaway JW, Conaway RC, Dvir A (2001) TFIIH action in transcription initiation and promoter escape requires distinct regions of downstream promoter DNA. *Proc Natl Acad Sci USA* 98(10):5544–5549.
- Fazal FM, Meng CA, Murakami K, Kornberg RD, Block SM (2015) Real-time observation of the initiation of RNA polymerase II transcription. *Nature* 525(7568):274–277.
- Murakami K, et al. (2015) Uncoupling promoter opening from start-site scanning. *Mol Cell* 59(1):133–138.
- He Y, Fang J, Taatjes DJ, Nogales E (2013) Structural visualization of key steps in human transcription initiation. *Nature* 495(7442):481–486.
- Kastner B, et al. (2008) GraFix: Sample preparation for single-particle electron cryomicroscopy. *Nat Methods* 5(1):53–55.
- Li X, et al. (2013) Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat Methods* 10(6):584–590.
- Yang Z, Fang J, Chittuluru J, Asturias FJ, Penczek PA (2012) Iterative stable alignment and clustering of 2D transmission electron microscope images. *Structure* 20(2):237–247.
- Lander GC, et al. (2009) Appion: An integrated, database-driven pipeline to facilitate EM image processing. *J Struct Biol* 166(1):95–102.
- 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. *Proc Natl Acad Sci USA* 95(16):9117–9122.
- Okuda M, et al. (2000) Structure of the central core domain of TFIIIB with a novel double-stranded DNA-binding surface. *EMBO J* 19(6):1346–1356.
- Meinhart A, Blobel J, Cramer P (2003) An extended winged helix domain in general transcription factor E11E alpha. *J Biol Chem* 278(48):48267–48274.
- Okuda M, Tanaka A, Hanaoka F, Ohkuma Y, Nishimura Y (2005) Structural insights into the asymmetric effects of zinc-ligand cysteine mutations in the novel zinc ribbon domain of human TFIIIB for transcription. *J Biochem* 138(4):443–449.
- Nikolov DB, et al. (1995) Crystal structure of a TFIIIB-TBP-TATA-element ternary complex. *Nature* 377(6545):119–128.
- Pettersen EF, et al. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25(13):1605–1612.
- Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66(Pt 4):486–501.
- Grünberg S, Warfield L, Hahn S (2012) Architecture of the RNA polymerase II pre-initiation complex and mechanism of ATP-dependent promoter opening. *Nat Struct Mol Biol* 19(8):788–796.
- 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(5523):1876–1882.