Architecture of an RNA Polymerase II Transcription Pre-Initiation Complex

Kenji Murakami, Hans Elmlund, Nir Kalisman, David A. Bushnell, Christopher M. Adams, Maia Azubel, Dominika Elmlund, Yael Levi-Kalisman, Xin Liu, Brian J. Gibbons, Michael Levitt, Roger D. Kornberg*

READ THE FULL ARTICLE ONLINE
http://dx.doi.org/10.1126/science.1238724

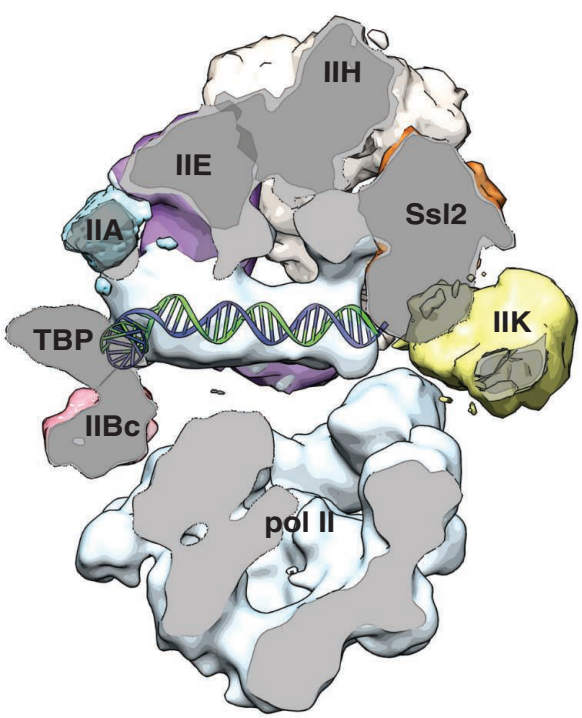
Cite this article as K. Murakami *et al.*, *Science* **342**, 1238724 (2013).
DOI: 10.1126/science.1238724

Introduction: RNA polymerase II (pol II) is capable of RNA synthesis but is unable to recognize a promoter or to initiate transcription. For these essential functions, a set of general transcription factors (GTFs)—termed TFIIB, -D, -E, -F, and -H—is required. The GTFs escort promoter DNA through the stages of recruitment to pol II, unwinding to create a transcription bubble, descent into the pol II cleft, and RNA synthesis to a length of 25 residues and transition to a stable elongating complex. The structural basis for these transactions is largely unknown. Only TFIIB has been solved by means of x-ray diffraction, in a complex with pol II. We report on the structure of a complete set of GTFs, assembled with pol II and promoter DNA in a 32-protein, 1.5 megaDalton "pre-initiation complex" (PIC), as revealed with cryo-electron microscopy (cryo-EM) and chemical cross-linking.

Methods: Three technical advances enabled the structural analysis of the PIC. First, a procedure was established for the preparation of a stable, abundant PIC. Both the homogeneity and functional activity of the purified PIC were demonstrated. Second, an algorithm was developed for alignment of cryo-EM images that requires no prior information (no "search model") and that can distinguish multiple conformational states. Last, a computational method was devised for determining the arrangement of protein subunits and domains within a cryo-EM density map from a pattern of chemical cross-linking.

Results: The density map of the PIC showed a pronounced division in two parts, one pol II and the other the GTFs. Promoter DNA followed a straight path, in contact with the GTFs but well separated from pol II, suspended above the active center cleft. Cross-linking and computational analysis led to a most probable arrangement of the GTFs, with IIB at the upstream end of the pol II cleft, followed by IIF, IIE, and IIH. The Ssl2 helicase subunit of IIH was located at the downstream end of the cleft.

Discussion: A principle of the PIC revealed by this work is the interaction of promoter DNA with the GTFs and not with pol II. The GTFs position the DNA above the pol II cleft, but interaction with pol II can only occur after melting of the DNA to enable bending for entry in the cleft. Contact of the DNA with the Ssl2 helicase in the PIC leads to melting (in the presence of adenosine triphosphatase). Cryo-EM by others, based on sequential assembly and analysis of partial complexes rather than of the complete PIC, did not show a separation between pol II and GTFs and revealed direct DNA–pol II interaction. The discrepancy calls attention to a role of the GTFs in preventing direct DNA-polymerase interaction.



A section through the cryo-EM structure of the complete PIC. Cut surfaces are shown in gray. Locations of densities due to pol II and the GTFs (TFIIA, TFIIB C-terminal domain, TBP subunit of TFIID, TFIIE, and TFIIH, including its helicase subunit Ssl2 and its kinase module TFIIK) are indicated. Density due to DNA is indicated by the superimposed double helix model. TFIIF is not seen in this section.

FIGURES IN THE FULL ARTICLE

SUPPLEMENTARY MATERIALS

The list of author affiliations is available in the full article online.
*Corresponding author. E-mail: kornberg@stanford.edu

Architecture of an RNA Polymerase II Transcription Pre-Initiation Complex

Kenji Murakami,[1]* Hans Elmlund,[1]* Nir Kalisman,[1]* David A. Bushnell,[1] Christopher M. Adams,[2] Maia Azubel,[1] Dominika Elmlund,[1] Yael Levi-Kalisman,[1]† Xin Liu,[1]‡ Brian J. Gibbons,[1] Michael Levitt,[1] Roger D. Kornberg[1]§

The protein density and arrangement of subunits of a complete, 32-protein, RNA polymerase II (pol II) transcription pre-initiation complex (PIC) were determined by means of cryogenic electron microscopy and a combination of chemical cross-linking and mass spectrometry. The PIC showed a marked division in two parts, one containing all the general transcription factors (GTFs) and the other pol II. Promoter DNA was associated only with the GTFs, suspended above the pol II cleft and not in contact with pol II. This structural principle of the PIC underlies its conversion to a transcriptionally active state; the PIC is poised for the formation of a transcription bubble and descent of the DNA into the pol II cleft.

Sixty proteins assemble in a 3-million Dalton complex at every RNA polymerase II (pol II) promoter, before every round of transcription (*1*, *2*). About half of these proteins, some 30 in number [the subunits of pol II and the general transcription factors (GTFs)], form a pre-initiation complex (PIC) that can recognize a minimal (TATA-box) promoter, select a transcription start site, and synthesize a nascent transcript. The remaining proteins are needed for recognition of an extended promoter (TAF complex) and for the regulation of transcription (Mediator complex) (*3*). Orchestration of the initiation process depends on the organization of components of the PIC. We report here on the three-dimensional (3D) arrangement of the 32 proteins of the PIC.

Biochemical studies have identified functions of several PIC proteins (*1*). TATA box–binding protein (TBP), a subunit of the general transcription factor TFIID, binds and bends TATA-box DNA (*4*). TFIIB brings the TBP-promoter DNA complex to pol II near the active center cleft (*5*). TFIIH, an 11-protein complex, has multiple catalytic activities, including helicase, Ssl2 (*6*), that melts promoter DNA to form the so-called "transcription bubble," and a protein kinase (*7*) whose action upon the C-terminal domain of pol II controls association with Mediator and other accessory proteins (*8*). Functional roles of additional GTFs TFIIA, TFIIE, and TFIIF are less well established.

Structural information on the PIC is incomplete. X-ray structures have been determined for pol II (*9*, *10*), pol II-DNA complexes (*11*), a pol II-TFIIB complex (*12*–*14*), and a TBP-TFIIB-DNA complex (*15*). Structures at much lower resolution have been determined by means of electron microscopy (EM) of negatively stained proteins for a pol II–TFIIF complex (*16*), for TFIIE (*17*), and for TFIIH (*18*–*20*). The advance reported here, structural analysis of a fully assembled PIC, was made possible by three technical developments. First, improved methods of preparation of the GTFs (*21*) and their assembly with pol II resulted in abundant, homogeneous PIC (*22*). Second, upon structural analysis by means of cryogenic electron microscopy (cryo-EM), it emerged that conformational heterogeneity was an impediment to image processing of the PIC. An algorithm was developed for classifying images on the basis of conformational state, enabling averaging and 3D reconstruction (*23*, *24*). Last, spatial proximities of proteins in the PIC were determined by means of cross-linking and mass spectroscopy (XL-MS) (*25*–*27*). A computational approach was devised for combining information from cryo-EM and XL-MS to arrive at a complete 3D map of the protein components of the PIC. Abbreviated results and discussion follow; additional results, discussion, and material and methods are presented in supplementary materials.

Cryo-EM and 3-D Reconstruction

Pol II and GTFs, isolated from the yeast *Saccharomyces cerevisiae*, were combined with a fragment of *HIS4* promoter DNA (–81/+1) and sedimented

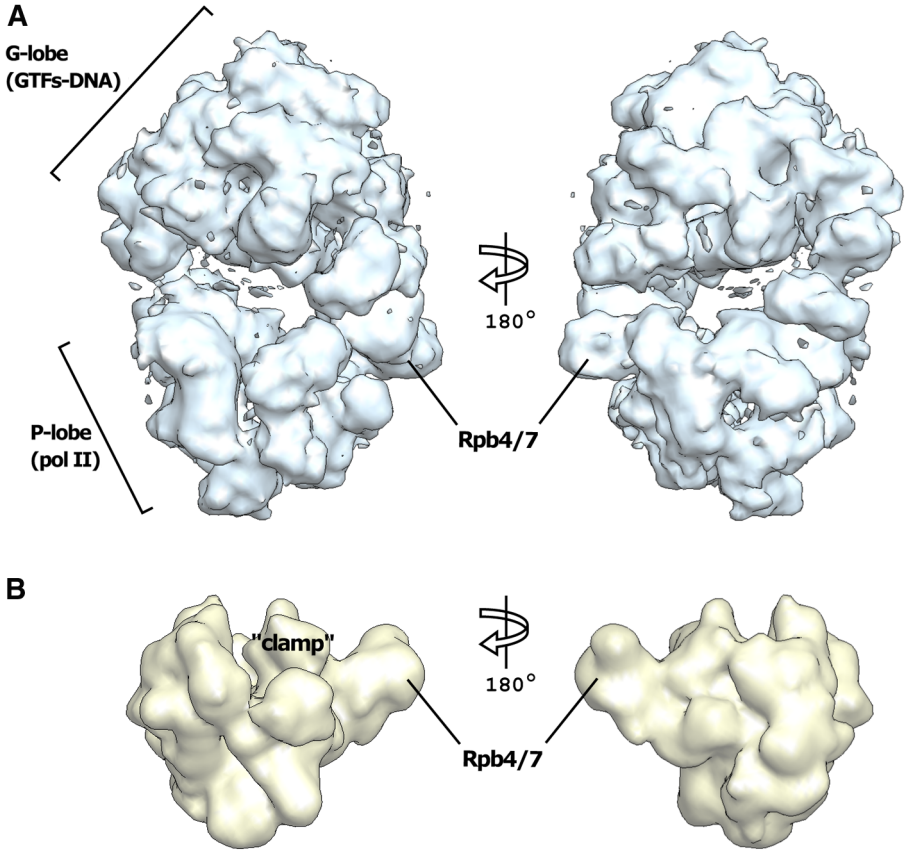


Fig. 1. Cryo-EM structure of the 32-protein PIC including TFIIS (PIC). (**A**) Front (left) and back (right) views of the most populated state of the PIC. (**B**) Low-pass filtered (20 Å) 12-subunit pol II (*45*) is viewed in the same orientations as in (A). The mobile "clamp" domain of pol II (*10*) and the dissociable subunits Rpb4/7 are indicated.

[1]Department of Structural Biology, Stanford University, Stanford, CA 94305, USA. [2]Stanford University Mass Spectrometry, Stanford University, Stanford, CA 94305, USA.

*These authors contributed equally to this work.
†Present address: Ilse Katz Institute for Nanoscale Science and Technology, Ben-Gurion University of the Negev, 84105 Beer-Sheva, Israel.
‡Present address: Cecil H. and Ida Green Center for Reproductive Biology Sciences, University of Texas Southwestern Medical Center, Dallas, TX 75390–8511, USA.
§Corresponding author. E-mail: kornberg@stanford.edu

in a glycerol gradient as described (fig. S1) (*22*). Two forms of the PIC were prepared in this way: one, denoted complete PIC, contained all GTFs (TFIIA, TFIIB, TBP, TFIIE, TFIIF, and TFIIH) and in addition the transcription elongation factor TFIIS, implicated through genetic analysis in the initiation of transcription in vivo (*28*) and with nuclear extract in vitro (*29*); a second form, denoted PIC-ΔTFIIK, was identical with the first, except for the removal of a three-subunit module of TFIIH termed TFIIK (within which resides the protein kinase mentioned above). Partial complexes, lacking one or more of the GTFs, are unstable and dissociate upon handling. It is uncertain whether partial complexes adopt defined structures. Only the complete PIC (or PIC-ΔTFIIK) exhibited a barrier to digestion by exonuclease III from the downstream end of the DNA. No barrier was observed when any one of the GTFs was omitted (fig. S1). Even the complete PIC tended to dissociate during specimen preparation for cryo-EM, and glutaraldehyde was included in the glycerol gradient (*30*) for stabilization. Images were acquired on a transmission electron microscope equipped with a field emission gun under low-dose conditions (10 to 15 $e^-/Å^2$) (fig. S2): 433,916 images of single particles of complete PIC and 305,160 images of PIC-ΔTFIIK.

Image processing was performed with SIMPLE (*24*), a program package for the analysis of asymmetrical and heterogeneous single particles that requires no prior knowledge of the structure, no search model, and therefore does not introduce model bias. SIMPLE addresses the problem of heterogeneity by means of a "Fourier common lines" approach (*23*). Analysis with SIMPLE revealed two predominant states in both forms of the PIC. After refinement, the resolution of the density maps was highest for complete PIC (16 Å by the Fourier shell criterion). The maps measured 240 × 170 × 150 Å. In both forms of the PIC, the maps were clearly divided in two parts, confirmed through automatic segmentation, and termed P-lobe and G-lobe (Fig. 1). An excellent fit of the pol II x-ray structure (contoured at 20 Å resolution) (Fig. 1B) to the P-lobe served to validate the segmentation and the reconstructions.

Segmentation of Electron Density: Location of GTFs and DNA

Manual fitting of the pol II structure to the P-lobe was optimized with computational refinement. A difference map calculated between the pol II structure and the P-lobe revealed one extra region of density in the P-lobe, adjacent to the Rpb4/7 subunits, which may be due to the C-terminal domain (CTD) of Rpb1 (fig. S3). The G-lobe showed similarity in size and shape to EM structures of negatively stained complete ("holo") TFIIH (Fig. 2, A and B) (*20*). On this basis, approximate locations of Rad3 and Ssl2 subunits of TFIIH were identified (color code to all subunits of GTFs is provided in Fig. 2D), and TFIIK was placed in close proximity to its substrate, the CTD. The location of TFIIK was confirmed by its removal from the PIC: The density attributed to TFIIK was absent from structures of PIC-ΔTFIIK (Fig. 2C).

The EM maps of the complete PIC also contained density attributable to TBP, TFIIA, TFIIB, and promoter DNA (Fig. 3). A crystallographic model (*13*, *14*) of a "minimal" PIC (pol II-TBP-TFIIB-DNA complex) could be docked to the EM density with only slight deviations in the DNA path. The DNA density contacted G-lobe density, 10 to 20 base pairs (bp) downstream of the TATA box and merged with Ssl2 density at the downstream end (Fig. 3, A and B). Because roughly cylindrical density for DNA could be seen between the upstream and downstream points of contact, it was essentially free in this region.

Removal of density due to TBP, TFIIA, TFIIB, TFIIH, and DNA from the G-lobe leaves residual density that may be attributed to TFIIE and TFIIF. The residual density was automatically segmented into regions of ~100 and 90 kD, which were assigned to TFIIE and TFIIF, respectively (Fig. 4A) on the basis of additional evidence from protein-protein cross-linking described below. Together, TFIIE and TFIIF would largely surround the DNA in the vicinity of the TATA box.

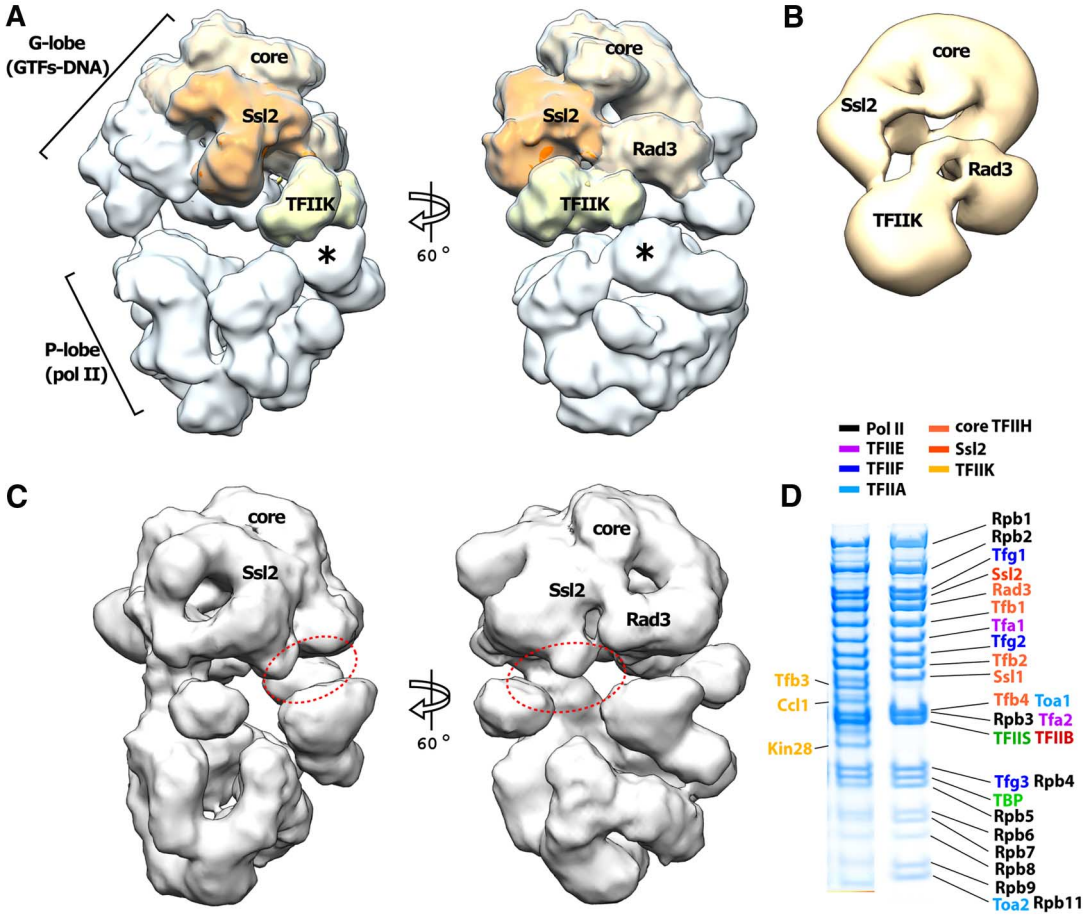


Fig. 2. Location of TFIIH in the PIC. (**A**) Front (left) and side (right) views of the PIC, with suggested assignments based on comparison with (B). Core and Rad3 are gold, Ssl2 is orange, and TFIIK is lime yellow. The extra density marked by the asterisk may be due to the CTD of Rpb1. (**B**) Molecular envelope from EM and 3D reconstruction of negatively stained TFIIH (*20*), viewed in the same orientation as (A), right. (**C**) Front (left) and side (right) views of the PIC lacking TFIIK (PIC-ΔTFIIK). Dashed ellipses indicate the location of the density, labeled TFIIK in (A), that is not observed. (**D**) SDS–polyacrylamide gel electrophoresis of peak glycerol gradient fractions of PIC (left) and PIC–ΔTFIIK (right). Rpb10, Rpb12, and Tfb5 are not seen because of their low molecular weights. The same color scheme is used in all figures.

XL-MS: Location of TFIIF

The proposed locations of the GTFs within the G-lobe, based on docking and automatic segmentation, were confirmed by means of XL-MS (*25*–*27*). For this purpose, the PIC was cross-linked with BS^3, a bifunctional amino group reagent. After protease digestion and mass spectrometry, assignments of cross-linked peptides to observed ion masses were scored for significance as described (*27*). A total of 109 intermolecular and 157 intramolecular cross-links of high significance (less than 1.5% false-positive rate) were identified (tables S1 and S2 and fig. S7). Validation came from 73 cross-links between residues separated by distances known from crystal structures of pol II-TFIIB (*14*), pol II-TFIIS (*31*), and TFIIA-TBP-DNA (*32*). The Cα-Cα distances between such residues were less than 30 Å for 70 cross-links, less than 34 Å for two cross-links, and 60 Å for the remaining one. These numbers are in accord with previous studies showing that the Cα-Cα distance of residues bridged by BS^3 are generally less than 30 Å, and in rare cases as much as 35Å (*25*, *27*). The one cross-link of 60 Å is likely a wrong assignment, which is not unexpected given the 1.5% false-positive rate.

Cross-links involving the Tfg1-Tfg2 dimerization and the Tfg2 winged-helix (WH) domains of TFIIF were sufficient to locate these domains in the EM map (Fig. 3B). The "flexible arm" of Tfg1 and the "insertion loop" of Tfg2 (*33*) formed a pattern of cross-links with residues in the Rpb2 subunit of pol II (K87, K344, K358, and K426) that placed the dimerization domain adjacent to the Rpb2 "protrusion" and "jaw-lobe" domains (Fig. 4C). The location of the insertion loop is supported by cross-links to the C-terminal region of Ssl2, which penetrates the downstream end of the pol II cleft, as shown by cross-linking of this region of Ssl2 to Rpb1 residues K1217, K1246, and K1262 (Fig. 4, A to C). The WH domain of Tfg2 also formed cross-links to the Toa2 subunit of TFIIA (K119), to TFIIB (K199) and to the Tfa2 subunit of TFIIE (Fig. 4B), that defined a specific location along the path of the DNA (table S4). The resulting model is in excellent agreement with previous FeBABE cleavage mapping of protein-promoter DNA interaction in complexes formed in yeast nuclear extract (*34*).

Combination of XL-MS and Cryo-EM: Locations of TFIIE and TFIIH

The two subunits of TFIIE and seven subunits of core TFIIH formed many cross-links with one another (fig. S7). To interpret this information in terms of the EM map, we represented each subunit by one or two spheres (fig. S12), depending on the mass of the subunit and the pattern of internal cross-links (a total of 12 spheres) (supplementary materials). Twelve locations for these spheres, spanning the EM density attributed to TFIIE and TFIIH (Fig. 5, A and B), were chosen by an objective procedure: The first location was at the point of highest EM density, the next location was at the point of next highest density that was more than 35 Å from the first, and so forth. There are 12 factorial (480 million) different models that assign the 12 spheres to the 12 locations, and we assessed them exhaustively. We first discarded models in which two spheres belonging to the same protein were more than 45 Å apart, reducing the number of models to half a million. We then evaluated the fit of a model to the pattern of cross-links on the basis of two measures: serious violation, defined by a pair of spheres located more than 65 Å apart in the model, for which a cross-link is nevertheless observed; and violation distance, defined as the excess over 40 Å between a pair of spheres in the model for which a cross-link is observed. These two measures were correlated over a wide range of values (Fig. 5B).

The model with the smallest sum of violation distances (Fig. 5A) was identical to the consensus of the 10 next best models at each location (Fig. 5D) and was the model returned most frequently by bootstrap analysis (fig. S13). The locations of TFIIE and TFIIH were the same as those determined with automatic segmentation of the EM density. The size and shape of the region assigned to core TFIIH were closely similar to those of a structure determined with 2D crystallography (Fig. 5C) (*18*). TFIIE extended across the surface of the G-lobe in proximity to the promoter DNA (Fig. 5A). Tfa1 and Tfa2 subunits of TFIIE ("Tfa1 N-term" and two WH domains of Tfa2) were cross-linked to each other immediately downstream of the TATA box (between DNA -54 and -44). "Tfa1 C-term" formed additional cross-links with all subunits of core TFIIH, reaching as far as Ssl1 (the N-terminal

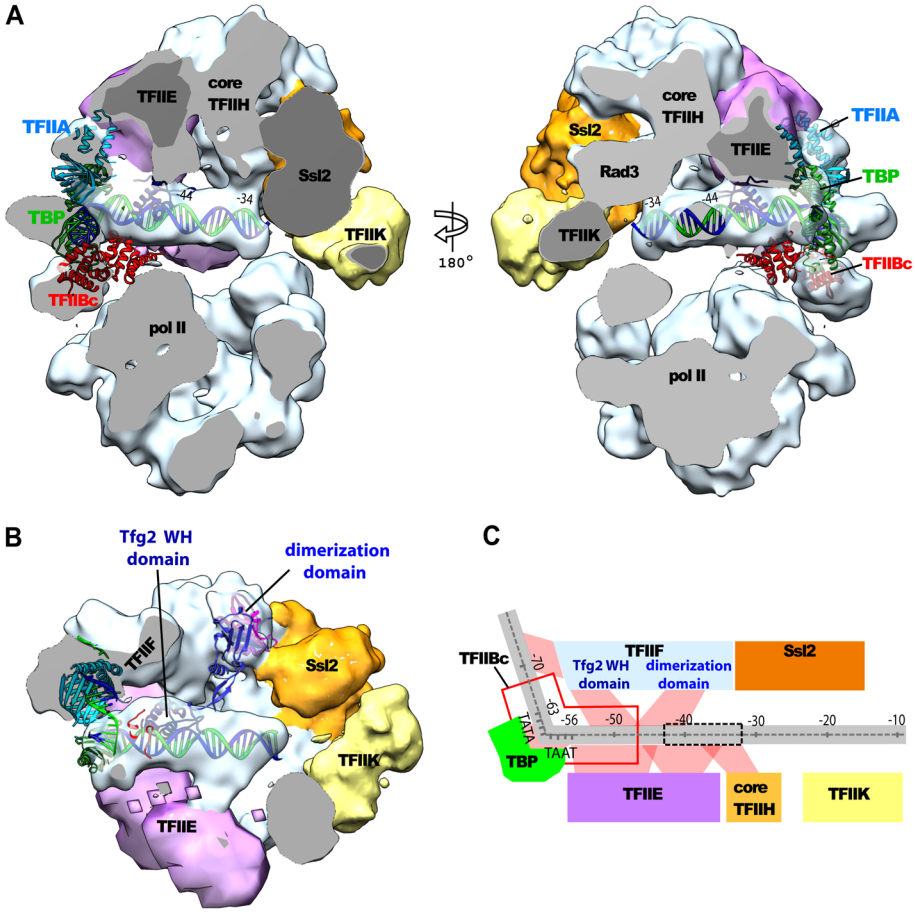


Fig. 3. Locations of general transcription factors and promoter DNA in the PIC. (**A**) The PIC structure, sectioned to reveal the DNA in the center (cut surfaces are gray). The structure on the left side of Fig. 1A was rotated 60° counterclockwise about the vertical axis (left) or 120° clockwise (right). Subunits of TFIIE and TFIIH (core TFIIH and its Rad3 component, Ssl2, and TFIIK) are indicated. Atomic models of the TBP (green)–$TFIIB_C$ (red)–TATA-box complex (*15*) and TFIIA (blue) (*32*) were fitted to the cryo-EM density and displayed as ribbon diagrams. The TATA box DNA was extended with straight B-form DNA with minor adjustments of the DNA path. (**B**) The underside of the G-lobe, viewed from bottom of (A), left, with the P-lobe removed. Atomic models of the Tfg1-Tfg2 dimerization domain (Tfg1, blue; Tfg2, magenta) and the winged helix (WH) domain of Tfg2 (dark blue) were located by means of protein-protein cross-linking and displayed as ribbon diagrams. (**C**) Schematic diagram of proximity relationships of promoter DNA (gray) and GTFs in the PIC. Positions in the DNA are numbered with respect to the first transcription start site of the *HIS4* promoter. The TATA box spans positions from –63 to –56. Transcription bubble formed upon initial promoter melting is indicated by the dashed box. Putative interactions are indicated with red shading.

half, denoted "Ssl1 N"). Ssl1 N formed cross-links to all subunits of core TFIIH, which is consistent with its assignment in the 2D crystal structure (*18*) to a central position, between Rad3 and the rest of the structure (Fig. 5C, right).

Discussion

This study has revealed a central principle of the PIC: the association of promoter DNA only with the GTFs and not with pol II. Promoter DNA is suspended above the pol II cleft, contacting three GTFs—TFIIB, TFIID (TBP subunit), and TFIIE—at the upstream end of the cleft (TATA box) and contacting TFIIH (Ssl2 helicase subunit) at the downstream end. In between, the DNA is free and available for action of the helicase, which untwists the DNA to introduce negative superhelical strain and thereby promote melting at a distance (*35*).

This principle of the PIC is a consequence of the rigidity of duplex DNA. The promoter duplex must follow a straight path, whereas bending through ~90° is required for binding in the pol II cleft (*11*). Only after melting can the DNA bend for entry in the cleft. Melting is thermally driven, induced by untwisting strain in the DNA above the cleft. A melted region is short-lived and must be captured by binding to pol II, which occurs rapidly enough because the DNA is positioned above the cleft. The GTFs therefore catalyze the formation of a stably melted region

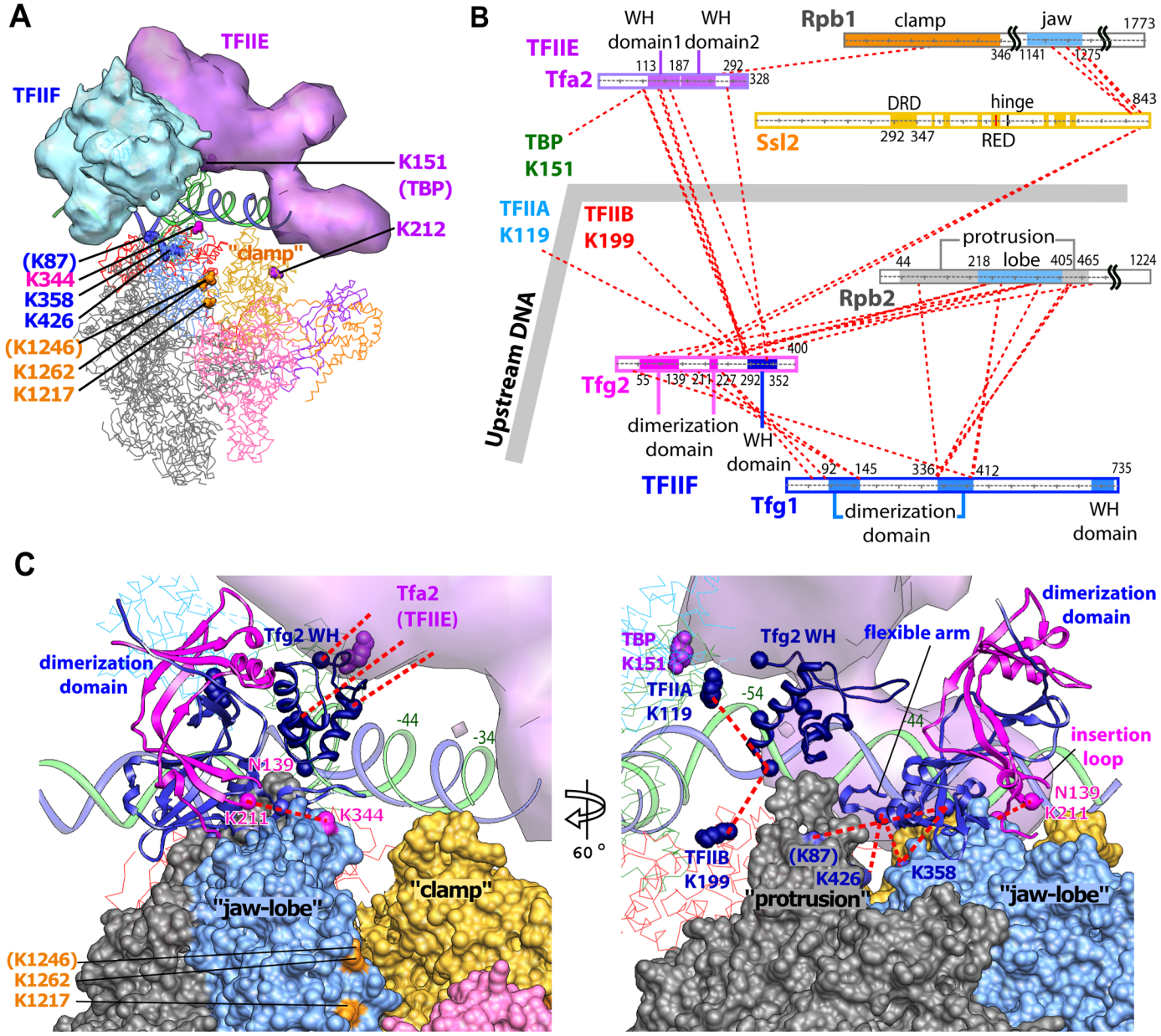
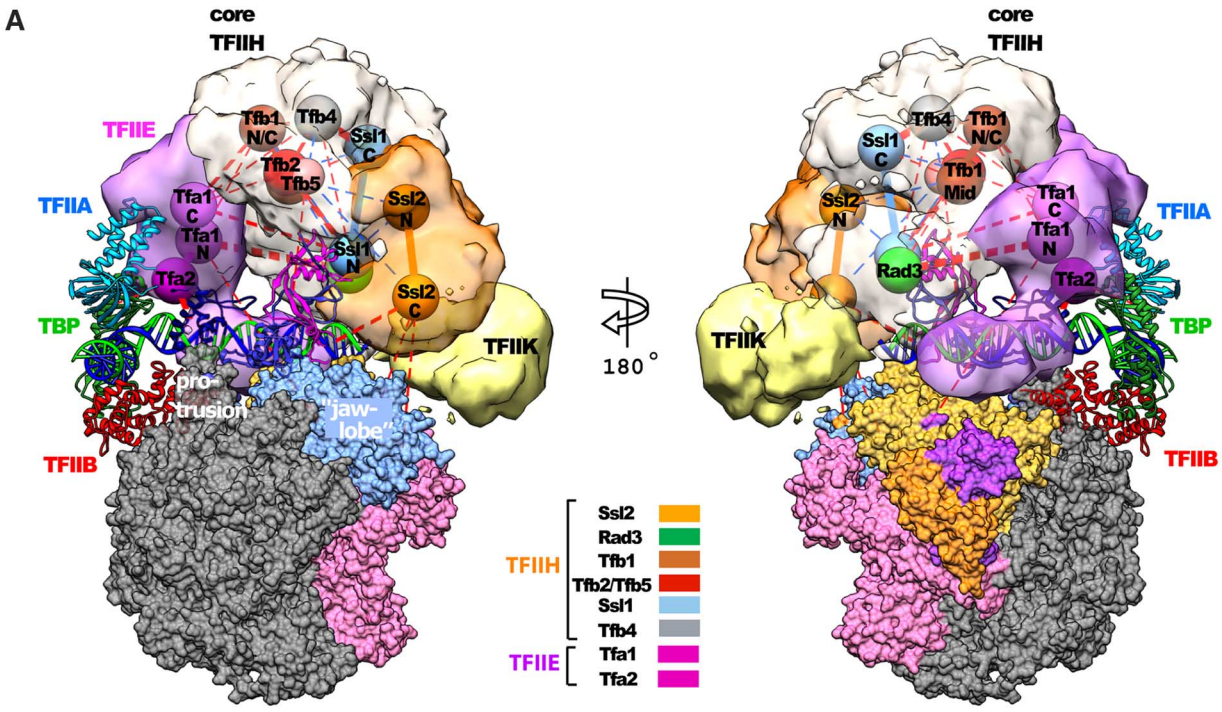
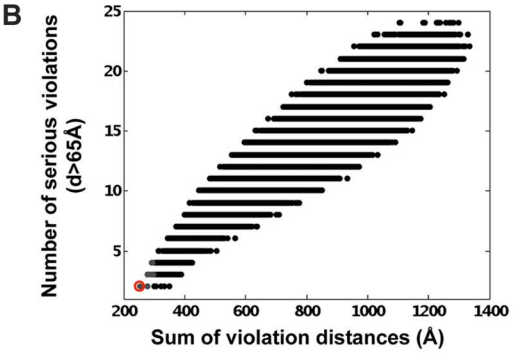
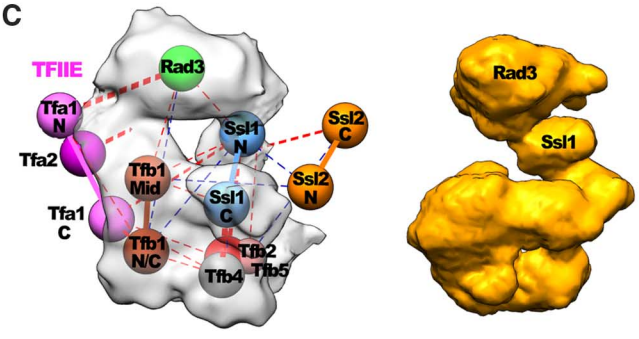


Fig. 4. Spatial restraints from XL-MS: domains of TFIIF. (**A**) Front view (same as Fig. 1A) of the PIC showing only the EM densities for TFIIF (light blue) and TFIIE (light purple). Pol II domains (jaw-lobe, light blue; clamp, yellow; core, gray; shelf, magenta), TFIIA (cyan), TFIIB (red), and TBP (green) are represented as backbone traces. Lysine residues of pol II and one of TBP that form cross-links to TFIIE and TFIIF are marked by van de Waals spheres and colored according to the subunit to which they are cross-linked (Tfg1, blue; Tfg2, magenta; TFIIE, purple; Ssl2, orange). If the cross-linked residue is absent from the model, then the closest structured residue (up to three residues away in the amino acid sequence) is shown in parentheses. (**B**) Schematic representation of cross-links (red dashed lines) involving Tfg1 and Tfg2, whose primary structures are depicted as boxes, with solid colors for regions of conserved folds. Only intersubunit cross-links are shown. (**C**) Close-up view of models of the Tfg1-Tfg2 dimerization domain (Tfg1, blue; Tfg2, magenta) and the WH domain of Tfg2 (dark blue) oriented on the basis of the cross-links indicated (red dashed lines). Pol II is shown in surface representation with the same colors as in (A). The Tfg2 yeast-specific insertion loop (residues 139 to 211) also cross-links to the C-terminal region of Ssl2, located near K1217, 1246, and 1262 of Rpb1.







D

Violation Distances	Violations	positions 1	2	3	4	5	6	7	8	9	10	11	12	Matches to #1	Rank
253.4	2	Tfa2	Tfa1-N	Tfa1-C	Rad3	Tfb1-N/C	Tfb1-Mid	Tfb4	Ssl1-N	Ssl1-C	Tfb2/Tfb5	Ssl2-N	Ssl2-C	12	1
278.2	2	Tfa2	Tfa1-N	Tfa1-C	Rad3	Tfb1-N/C	Tfb1-Mid	Tfb4	Ssl1-C	Ssl1-N	Tfb2/Tfb5	Ssl2-N	Ssl2-C	10	2
278.9	3	Tfa1-N	Tfa2	Tfa1-C	Rad3	Tfb1-N/C	Tfb1-Mid	Tfb4	Ssl1-N	Ssl1-C	Tfb2/Tfb5	Ssl2-N	Ssl2-C	10	3
284.7	3	Tfa2	Rad3	Tfa1-N	Ssl1-C	Tfa1-C	Tfb1-Mid	Tfb4	Ssl1-N	Tfb1-N/C	Tfb2/Tfb5	Ssl2-N	Ssl2-C	7	4
285.0	3	Tfa2	Tfa1-N	Rad3	Tfb1-N/C	Ssl1-N	Tfa1-C	Ssl1-C	Tfb1-Mid	Tfb4	Tfb2/Tfb5	Ssl2-N	Ssl2-C	5	5
290.3	3	Tfa2	Tfa1-N	Tfa1-C	Rad3	Tfb1-N/C	Tfb1-Mid	Ssl1-C	Tfb2/Tfb5	Ssl1-N	Tfb4	Ssl2-N	Ssl2-C	8	6
290.4	3	Tfa2	Tfa1-N	Tfa1-C	Rad3	Tfb1-N/C	Tfb1-Mid	Tfb4	Ssl2-N	Ssl1-N	Tfb2/Tfb5	Ssl1-C	Ssl2-C	9	7
291.7	4	Tfa2	Tfa1-N	Rad3	Tfb1-N/C	Ssl1-N	Tfa1-C	Tfb4	Tfb1-Mid	Ssl1-C	Tfb2/Tfb5	Ssl2-N	Ssl2-C	7	8
294.9	4	Tfa2	Tfa1-N	Rad3	Tfb1-N/C	Ssl1-N	Tfa1-C	Tfb2/Tfb5	Tfb1-Mid	Tfb4	Ssl1-C	Ssl2-N	Ssl2-C	4	9
296.0	3	Tfa2	Tfa1-N	Tfa1-C	Rad3	Tfb1-N/C	Tfb1-Mid	Tfb2/Tfb5	Ssl1-N	Ssl1-C	Tfb4	Ssl2-N	Ssl2-C	10	10
297.6	3	Tfa2	Tfa1-N	Rad3	Tfb1-Mid	Ssl1-N	Tfa1-C	Tfb4	Tfb1-N/C	Ssl1-C	Tfb2/Tfb5	Ssl2-N	Ssl2-C	7	11
Consensus		Tfa2	Tfa1-N	Tfa1-C	Rad3	Tfb1-N/C	Tfb1-Mid	Tfb4	Ssl1-N	Ssl1-C	Tfb2/Tfb5	Ssl2-N	Ssl2-C	12	
Matches to #1		10	9	6	6	6	7	7	4	5	8	10	11		

Fig. 5. Combination of XL-MS and cryo-EM: TFIIE and TFIIH. (**A**) Side views of PIC showing EM densities attributed to TFIIE (light purple) and TFIIH (core TFIIH, gray; Ssl2, orange; TFIIK, light yellow). Spheres for TFIIE, core TFIIH, and Ssl2 subunits are labeled according to the model that best fits the XL-MS data. Solid lines connect spheres belonging to the same subunit. Red dashed lines indicate intersubunit cross-links in the PIC, with thicknesses proportional to the number of cross-links observed. Blue dashed lines indicate cross-links in holo-TFIIH only. The EM density for TFIIF is omitted for clarity. Other elements of the PIC are represented as in Fig. 4C. (**B**) Fit of various models of the subunit locations (scatter) to XL-MS data. Two measures of fit are plotted for each model: the number of cross-linked spheres that are more than 65 Å apart (*y* axis) and the sum of distances in excess of 40 Å between cross-linked spheres (*x* axis). The best-fitting model (red circle) is shown in (A). (**C**) Comparison of the density for core TFIIH (left) from (A) with the reconstructed volume of core TFIIH from EM of 2D crystals in stain (right) (*18*). Locations of TFIIE (three purple spheres) and Ssl2 (two orange spheres) from (A) are shown on the left as well. (**D**) Listing of the 11 best fitting models [gray scatter in (B)], detailing for each the assignment of the subunits to the 12 fixed positions in the electron density. Each row represents a different model, with the best-fitting model (A) listed at the top. The most frequently occurring subunit at each position (the consensus) is identical to that of the best-fitting model. The strength of the consensus (bottom row) is indicative of the confidence of subunit assignment.

(transcription bubble) in two ways, by the introduction of untwisting strain (by the helicase) and by positioning promoter DNA.

Untwisting strain is distributed throughout the DNA above the pol II cleft, so melting may occur at any point, but only a melted region adjacent to TFIIB is stabilized by binding to pol II. The reason is again the rigidity of duplex DNA, and the requirement for a sharp bend adjacent to TFIIB to penetrate the pol II cleft. A single strand of DNA must extend from the point of contact with TFIIB, ~13 bp downstream of the TATA box (*36*, *37*), through the binding site for the transcription bubble in pol II. TFIIB may also interact with the single strand to stabilize the bubble (*14*).

Our conclusions are based on results from both cryo-EM and XL-MS, which served to validate one another: Segmentation and labeling of electron density, based on fitting pol II and other known structures, was consistent with all but three of 266 cross-links observed. Our PIC structure is also consistent with partial structural information from x-ray crystallography [pol II–TFIIB (*12*–*14*), pol II–TFIIS (*31*), TFIIA–TBP–TFIIB–DNA (*15*, *32*), and Tfb2–Tfb5 (*38*)], from nuclear magnetic resonance [Tfb1–Tfa1 (*39*) and Tfa2–DNA (*40*)], and from EM [core and holo TFIIH (*18*, *20*)]. This consistency provides cross-validation, both supporting our PIC structure and establishing the relevance of the partial structural information. Further consistency with the results of FeBABE cleavage mapping of complexes formed in yeast nuclear extract (*34*) was mentioned above; the locations of proteins along the DNA in our PIC structure and those determined with FeBABE cleavage differ by no more than 5 bp. Our PIC structure also agrees with results of protein-DNA cross-linking in a reconstituted human transcription system (*35*); positions of TFIIE and TFIIH differ between the two studies by ~20 and 10 bp. The location of Ssl2 in our structure, ~30 bp downstream from the TATA box (Fig. 3, B and C), supports the proposal, made on the basis of previous DNA-protein cross-linking analysis, that helicase action torques the DNA to introduce untwisting strain and thereby to promote melting at a distance (*35*).

While this manuscript was in preparation, Nogales and coworkers reported EM structures of human PICs (*41*). The Nogales structures were produced from presumptive partial complexes through the alignment of EM images to a structure of pol II as a search model, so they unavoidably include the pol II structure. Beyond pol II and two GTF polypeptides, TBP and TFIIB, there is little resemblance between the Nogales structures and ours (Fig. 6, A and B, and fig. S14). The Nogales structures have no G-lobe (Fig. 6B); contain very little density for TFIIE (30% of expected) and TFIIF (20%); are inconsistent with the majority of the intermolecular cross-links involving TFIIE, TFIIF, and TFIIH that we observed; and show a very different DNA path from our structure (Fig. 7). In our view, the Nogales structures do not reveal a complete PIC, but only a complex of pol II with TBP, TFIIB, and DNA, for two reasons: First, partial complexes are unstable, so TFIIE and TFIIF failed to bind or were lost; and second, because image processing was performed with the pol II structure as a search model, conformational flexibility was neglected, and only proteins rigidly bound to pol II were observed. Differences from our complete PIC structure and the reasons for these differences are described in more detail in the supplementary materials.

The difference in DNA path between the Nogales structures and ours explains a difference

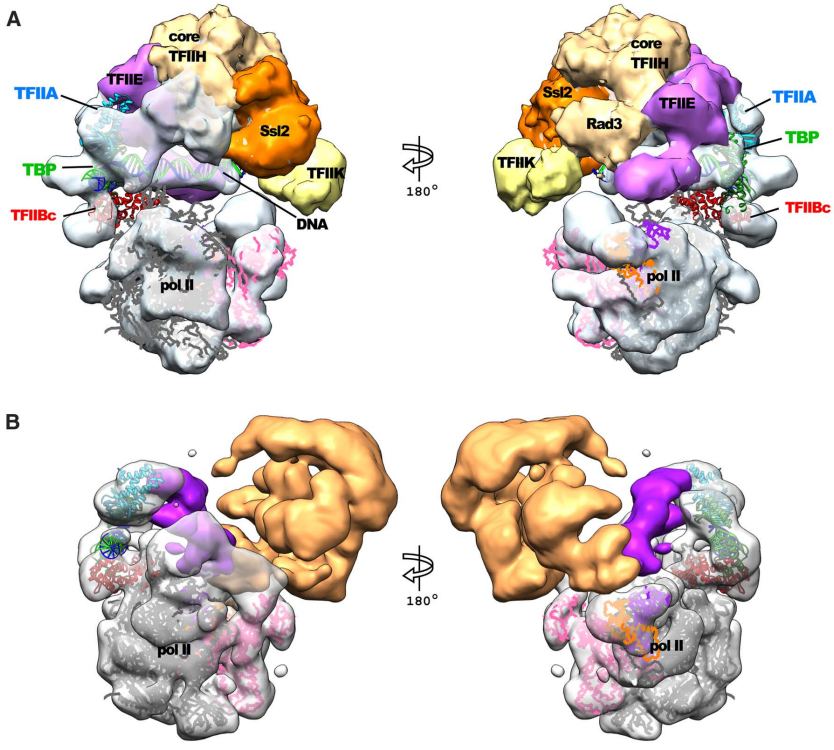


Fig. 6. Comparison of reconstructed volume of the PIC in this work to that of negatively stained human PIC. (**A**) Side views of the PIC from this work. Rotated by 60° (left) or −120° (right) relative to front view in Fig. 1A. (**B**) The reconstruction from negatively stained human PIC (*41*) is viewed in the same pol II orientation as in (A). The extra density due to TFIIH (light brown) is only weakly anchored to the remainder of the PIC. The center of mass of TFIIH is displaced by ~90 Å compared with the structure in (A). TFIIE (purple) is dissociated from TBP and TFIIH. Density due to DNA is absent.

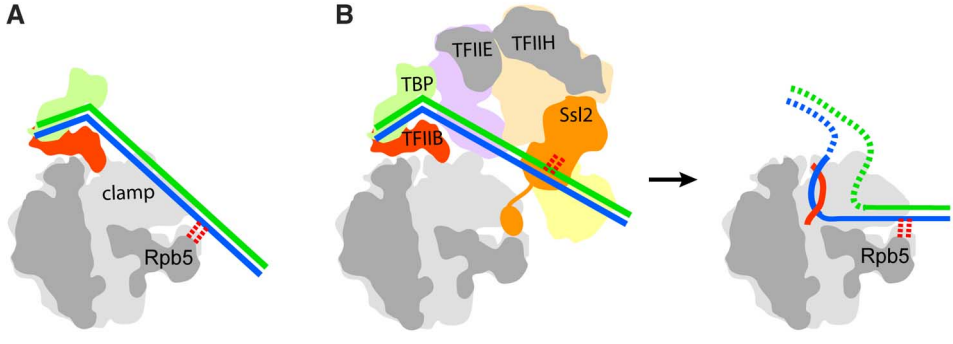


Fig. 7. Comparison of DNA paths within PIC structures. (**A**) Cutaway view in schematic representation of cryo-EM structure of PIC lacking TFIIE and TFIIH (*41*). TFIIF is omitted for clarity. DNA is blue and green. The downstream DNA contacts (red dotted lines) the tip of Rpb5. (**B**) (Left) Cutaway view in schematic representation of the complete PIC in this study. TFIIF is omitted for clarity. The C-terminal region of Ssl2 (orange appendage at the bottom of Ssl2) binds residues within the pol II cleft, preventing entry of DNA, which is instead suspended above the cleft, where it interacts with Ssl2 (red dotted lines). (Right) Cutaway view for the transcribing complex based on the x-ray structure (*11*), and contact with Rpb5 in the cleft is indicated (red dotted lines). Nascent RNA is red.

between two reports of protein-DNA cross-linking of pol II–GTF-DNA complexes and suggests a role for the C-terminal domain of Ssl2; a high-frequency cross-link of Rpb5 to DNA downstream of the TATA box was identified for a PIC lacking TFIIH (*42*, *43*), whereas no such cross-link was found for a complete PIC, but instead a cross-link of Ssl2 was observed at the same position (*34*). The trajectory of the DNA in the Nogales structures, penetrating the pol II cleft at the location of Rpb5, is thus explained by the absence of TFIIE and TFIIH, whereas the location of the DNA in our structure is due to the presence of TFIIH. The position of DNA in the Nogales structures would clash with the location of the C-terminal domain of Ssl2 in the cleft in our structure. We suggest that the site bound by DNA in the Nogales structure is blocked by the C-terminal domain of Ssl2; only after promoter melting and entry of the template strand at the upstream end of the pol II cleft does DNA displace the C-terminal domain of Ssl2 and occupy the downstream end of the cleft (Fig. 7B). Region 1.1 of σ70 performs a similar role, binding in the cleft of bacterial RNA polymerase, preventing nonspecific DNA interaction, and dissociating upon open complex formation (*44*).

References and Notes

1. R. C. Conaway, J. W. Conaway, General initiation factors for RNA polymerase II. *Annu. Rev. Biochem.* **62**, 161–190 (1993). doi: 10.1146/annurev.bi.62.070193.001113; pmid: 8352588
2. R. D. Kornberg, The molecular basis of eukaryotic transcription. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12955–12961 (2007). doi: 10.1073/pnas.0704138104; pmid: 17670940
3. R. D. Kornberg, Mediator and the mechanism of transcriptional activation. *Trends Biochem. Sci.* **30**, 235–239 (2005). doi: 10.1016/j.tibs.2005.03.011; pmid: 15896740
4. S. K. Burley, R. G. Roeder, Biochemistry and structural biology of transcription factor IID (TFIID). *Annu. Rev. Biochem.* **65**, 769–799 (1996). doi: 10.1146/annurev.bi.65.070196.004005; pmid: 8811195
5. S. Buratowski, S. Hahn, L. Guarente, P. A. Sharp, Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell* **56**, 549–561 (1989). doi: 10.1016/0092-8674(89)90578-3; pmid: 2917366
6. S. N. Guzder, P. Sung, V. Bailly, L. Prakash, S. Prakash, RAD25 is a DNA helicase required for DNA repair and RNA polymerase II transcription. *Nature* **369**, 578–581 (1994). doi: 10.1038/369578a0; pmid: 8202161
7. W. J. Feaver, O. Gileadi, Y. Li, R. D. Kornberg, CTD kinase associated with yeast RNA polymerase II initiation factor b. *Cell* **67**, 1223–1230 (1991). doi: 10.1016/0092-8674(91)90298-D; pmid: 1836979
8. Y. J. Kim, S. Björklund, Y. Li, M. H. Sayre, R. D. Kornberg, A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* **77**, 599–608 (1994). doi: 10.1016/0092-8674(94)90221-6; pmid: 8187178
9. P. Cramer *et al.*, Architecture of RNA polymerase II and implications for the transcription mechanism. *Science* **288**, 640–649 (2000). doi: 10.1126/science.288.5466.640; pmid: 10784442
10. P. Cramer, D. A. Bushnell, R. D. Kornberg, Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* **292**, 1863–1876 (2001). doi: 10.1126/science.1059493; pmid: 11313498
11. A. L. Gnatt, P. Cramer, J. Fu, D. A. Bushnell, R. D. Kornberg, Structural basis of transcription: An RNA polymerase II elongation complex at 3.3 A resolution. *Science* **292**, 1876–1882 (2001). doi: 10.1126/science.1059495; pmid: 11313499
12. D. A. Bushnell, K. D. Westover, R. E. Davis, R. D. Kornberg, Structural basis of transcription: An RNA polymerase II-TFIIB cocrystal at 4.5 Angstroms. *Science* **303**, 983–988 (2004). doi: 10.1126/science.1090838; pmid: 14963322
13. D. Kostrewa *et al.*, RNA polymerase II-TFIIB structure and mechanism of transcription initiation. *Nature* **462**, 323–330 (2009). doi: 10.1038/nature08548; pmid: 19820686
14. X. Liu, D. A. Bushnell, D. Wang, G. Calero, R. D. Kornberg, Structure of an RNA polymerase II-TFIIB complex and the transcription initiation mechanism. *Science* **327**, 206–209 (2010). doi: 10.1126/science.1182015; pmid: 19965383
15. D. B. Nikolov *et al.*, Crystal structure of a TFIIB-TBP-TATA-element ternary complex. *Nature* **377**, 119–128 (1995). doi: 10.1038/377119a0; pmid: 7675079
16. W. H. Chung *et al.*, RNA polymerase II/TFIIF structure and conserved organization of the initiation complex. *Mol. Cell* **12**, 1003–1013 (2003). doi: 10.1016/S1097-2765(03)00387-3; pmid: 14580350
17. A. Jawhari *et al.*, Structure and oligomeric state of human transcription factor TFIIE. *EMBO Rep.* **7**, 500–505 (2006). pmid: 16547462
18. W. H. Chang, R. D. Kornberg, Electron crystal structure of the transcription factor and DNA repair complex, core TFIIH. *Cell* **102**, 609–613 (2000). doi: 10.1016/S0092-8674(00)00083-0; pmid: 11007479
19. P. Schultz *et al.*, Molecular structure of human TFIIH. *Cell* **102**, 599–607 (2000). doi: 10.1016/S0092-8674(00)00082-9; pmid: 11007478
20. B. J. Gibbons *et al.*, Subunit architecture of general transcription factor TFIIH. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 1949–1954 (2012). doi: 10.1073/pnas.1105266109; pmid: 22308316
21. K. Murakami *et al.*, Tfb6, a previously unidentified subunit of the general transcription factor TFIIH, facilitates dissociation of Ssl2 helicase after transcription initiation. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 4816–4821 (2012). doi: 10.1073/pnas.1201448109; pmid: 22411836
22. K. Murakami *et al.*, Formation and fate of a complete 31-protein RNA polymerase II transcription preinitiation complex. *J. Biol. Chem.* **288**, 6325–6332 (2013). doi: 10.1074/jbc.M112.433623; pmid: 23303183
23. D. Elmlund, R. Davis, H. Elmlund, Ab initio structure determination from electron microscopic images of single molecules coexisting in different functional states. *Structure* **18**, 777–786 (2010). doi: 10.1016/j.str.2010.06.001; pmid: 20637414
24. D. Elmlund, H. Elmlund, SIMPLE: Software for ab initio reconstruction of heterogeneous single-particles. *J. Struct. Biol.* **180**, 420–427 (2012). doi: 10.1016/j.jsb.2012.07.010; pmid: 22902564
25. Z. A. Chen *et al.*, Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry. *EMBO J.* **29**, 717–726 (2010). doi: 10.1038/emboj.2009.401; pmid: 20094031
26. K. Lasker *et al.*, Molecular architecture of the 26S proteasome holocomplex determined by an integrative approach. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 1380–1387 (2012). doi: 10.1073/pnas.1120559109; pmid: 22307589
27. N. Kalisman, C. M. Adams, M. Levitt, Subunit order of eukaryotic TRiC/CCT chaperonin by cross-linking, mass spectrometry, and combinatorial homology modeling. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 2884–2889 (2012). doi: 10.1073/pnas.1119472109; pmid: 22308438
28. D. M. Prather, E. Larschan, F. Winston, Evidence that the elongation factor TFIIS plays a role in transcription initiation at GAL1 in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **25**, 2650–2659 (2005). doi: 10.1128/MCB.25.7.2650-2659.2005; pmid: 15767671
29. B. Kim *et al.*, The transcription elongation factor TFIIS is a component of RNA polymerase II preinitiation complexes. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 16068–16073 (2007). doi: 10.1073/pnas.0704573104; pmid: 17913884
30. B. Kastner *et al.*, GraFix: Sample preparation for single-particle electron cryomicroscopy. *Nat. Methods* **5**, 53–55 (2008). doi: 10.1038/nmeth1139; pmid: 18157137
31. H. Kettenberger, K. J. Armache, P. Cramer, Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Mol. Cell* **16**, 955–965 (2004). doi: 10.1016/j.molcel.2004.11.040; pmid: 15610738
32. S. Tan, Y. Hunziker, D. F. Sargent, T. J. Richmond, Crystal structure of a yeast TFIIA/TBP/DNA complex. *Nature* **381**, 127–134 (1996). doi: 10.1038/381127a0; pmid: 8610010
33. F. Gaiser, S. Tan, T. J. Richmond, Novel dimerization fold of RAP30/RAP74 in human TFIIF at 1.7 A resolution. *J. Mol. Biol.* **302**, 1119–1127 (2000). doi: 10.1006/jmbi.2000.4110; pmid: 11183778
34. G. Miller, S. Hahn, A DNA-tethered cleavage probe reveals the path for promoter DNA in the yeast preinitiation complex. *Nat. Struct. Mol. Biol.* **13**, 603–610 (2006). doi: 10.1038/nsmb1117; pmid: 16819517
35. T. K. Kim, R. H. Ebright, D. Reinberg, Mechanism of ATP-dependent promoter melting by transcription factor IIH. *Science* **288**, 1418–1421 (2000). doi: 10.1126/science.288.5470.1418; pmid: 10827951
36. F. C. Holstege, U. Fiedler, H. T. Timmers, Three transitions in the RNA polymerase II transcription complex during initiation. *EMBO J.* **16**, 7468–7480 (1997). doi: 10.1093/emboj/16.24.7468; pmid: 9405375
37. M. Pal, A. S. Ponticelli, D. S. Luse, The role of the transcription bubble and TFIIB in promoter clearance by RNA polymerase II. *Mol. Cell* **19**, 101–110 (2005). doi: 10.1016/j.molcel.2005.05.024; pmid: 15989968
38. D. E. Kainov, M. Vitorino, J. Cavarelli, A. Poterszman, J. M. Egly, Structural basis for group A trichothiodystrophy. *Nat. Struct. Mol. Biol.* **15**, 980–984 (2008). doi: 10.1038/nsmb.1478; pmid: 19172752
39. M. Okuda *et al.*, Structural insight into the TFIIE-TFIIH interaction: TFIIE and p53 share the binding region on TFIIH. *EMBO J.* **27**, 1161–1171 (2008). doi: 10.1038/emboj.2008.47; pmid: 18354501
40. M. Okuda *et al.*, Structure of the central core domain of TFIIEbeta with a novel double-stranded DNA-binding surface. *EMBO J.* **19**, 1346–1356 (2000). doi: 10.1093/emboj/19.6.1346; pmid: 10716934
41. Y. He, J. Fang, D. J. Taatjes, E. Nogales, Structural visualization of key steps in human transcription initiation. *Nature* **495**, 481–486 (2013). doi: 10.1038/nature11991; pmid: 23446344
42. T. K. Kim *et al.*, Trajectory of DNA in the RNA polymerase II transcription preinitiation complex. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 12268–12273 (1997). doi: 10.1073/pnas.94.23.12268; pmid: 9356438
43. D. Forget, M. F. Langelier, C. Thérien, V. Trinh, B. Coulombe, Photo-cross-linking of a purified preinitiation complex reveals central roles for the RNA polymerase II mobile clamp and TFIIE in initiation mechanisms. *Mol. Cell. Biol.* **24**, 1122–1131 (2004). doi: 10.1128/MCB.24.3.1122-1131.2004; pmid: 14729958
44. V. Mekler *et al.*, Structural organization of bacterial RNA polymerase holoenzyme and the RNA polymerase-promoter open complex. *Cell* **108**, 599–614 (2002). doi: 10.1016/S0092-8674(02)00667-0; pmid: 11893332
45. D. A. Bushnell, R. D. Kornberg, Complete, 12-subunit RNA polymerase II at 4.1-A resolution: Implications for the initiation of transcription. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 6969–6973 (2003). doi: 10.1073/pnas.1130601100; pmid: 12746498

Acknowledgments: This research was supported by NIH grants GM49885 and AI21144 to R.D.K. and GM063817 to M.L. The EM map has been deposited in the EMDataBank under accession code EMD-2394.

Supplementary Materials

www.sciencemag.org/content/342/6159/1238724/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S14
Tables S4 and S5
References (*46*–*63*)
Tables S1 to S3
Movies S1 to S3
Protein Data Bank File for the PIC Model

4 April 2013; accepted 5 September 2013
10.1126/science.1238724

Science

Architecture of an RNA Polymerase II Transcription Pre-Initiation Complex

Kenji Murakami, Hans Elmlund, Nir Kalisman, David A. Bushnell, Christopher M. Adams, Maia Azubel, Dominika Elmlund, Yael Levi-Kalisman, Xin Liu, Brian J. Gibbons, Michael Levitt and Roger D. Kornberg

Science **342** (6159), 1238724.
DOI: 10.1126/science.1238724originally published online September 26, 2013

Pre-Initiation Complex in 3D

The regulation of gene expression is critical for almost every aspect of biology. Transcription—generating an RNA copy of a gene —requires the assembly of a large pre-initiation complex (PIC) at every RNA polymerase II (pol II) promoter. Roughly 32 proteins—the subunits of pol II and the general transcription factors—form a PIC that can recognize a minimal TATA-box promoter, select a transcription start site, and synthesize a nascent transcript. **Murakami *et al.*** (p. 10.1126/science.1238724, published online 26 September; see the Perspective by **Malik and Roeder**) determined the three-dimensional map of the *Saccharomyces cerevisiae* 30-subunit PIC using cryo-electron microscopy. The saddle-shaped TATA binding protein, the boot-shaped transcription factor IIA (TFIIA), and promoter DNA ~27 bp downstream of the TATA-box could all be seen. Cross-linking and mass spectrometry was used to determine the spatial proximity of the 30 subunits, revealing that the PIC forms two lobes with TFIIF forming a bridge between them.