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, Michael Levitt,*

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, 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 demonstrationed. 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.

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

Published by AAAS
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

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 Dal-ton 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 TFIIID, binds and bends TATA-box DNA (4). TFIIIB 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 TFIIIF 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 spectrometry (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...
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, TFIIIB, TBP, TFIIE, TFIIF, and TFIIH) and in addition the transcription elongation factor TFIIIS, 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⁻/Å²) (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 domains of TFIIA (Fig. 4A). Density attributable to the 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, TFIIIB, and promoter DNA (Fig. 3). A crystallographic model (13, 14) of a “minimal” PIC (pol II-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, TFIIIB, TFIIH, and DNA from the G-lobe leaves residual density that may be attributed to TFIIE and TFIIF. The residual density was automatically segregated 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.
XL-M5: 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-M5 (25–27). For this purpose, the PIC was cross-linked with BS3, 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 intracellular 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 TFIIBA-TBP-DNA (32). The Ca-Ca 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 Ca-Ca distance of residues bridged by BS3 are generally less than 30 Å, and in rare cases as much as 35 Å (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 TFIIF (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-M5 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 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 TFIE (“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 TFIIIE and TFIIH (core TFIIH and its Rad3 component, Ssl2, and TFIIB) are indicated. Atomic models of the TBP (green)–TFIIA (red)–TATA-box complex (15) and TFIIB (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.](http://science.sciencemag.org/content/342/6158/1238724/F3.large.jpg)
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, TFID (TBP subunit), and TFIE— 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 TFIE (light purple). Pol II domains (jaw-lobe, light blue; clamp, yellow; core, gray; shelf, magenta); TFIA (cyan), TFIIF (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; TFIE, 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.
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 sum of distances in excess of 40 Å between cross-linked spheres (x axis) and the sum of distances in excess of 65 Å apart (y axis) and 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 model 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.
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 TFII B 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 TFII B to penetrate the pol II cleft. A single strand of DNA must extend from the point of contact with TFII B, ~13 bp downstream of the TATA box (36, 37), through the binding site for the transcription bubble in pol II. TFII B 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–TFII B (12–14), pol II–TFII H (31), TFII A–TBP–TFII B–DNA (15, 32), and Tbk2–Tbk5 (38)], from nuclear magnetic resonance [Tb1–Tfa1 (39) and Tfa2–DNA (40)], and from EM [core and holo TFII H (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 FeBEB E 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 FeBEB E 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 TFII E and TFII H 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, Nagules and coworkers reported EM structures of human PICs (41). The Nagules 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 TFII B, there is little resemblance between the Nagules structures and ours (Fig. 6, A and B, and fig. S14). The Nagules structures have no G-lobe (Fig. 6B); contain very little density for TFII E (30% of expected) and TFII H (20%); are inconsistent with the majority of the intermolecular cross-links involving TFII E, TFII F, and TFII H that we observed; and show a very different DNA path from our structure (Fig. 7). In our view, the Nagules structures do not reveal a complete PIC, but only a complex of pol II with TBP, TFII B, and DNA, for two reasons: First, partial complexes are unstable, so TFII E and TFII H 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 Nagules structures and ours explains a difference
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 Rp5 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 cleat at the location of Rp5, is thus explained by the absence of TFIIH and TFIIJ, 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 cleat in our structure. We suggest that the site bound by DNA in the Nogales structure is blocked by the C-terminal domain of Ssl2; a high-frequency binding in the cleft of bacterial RNA polymerase, of Ssl2 was observed at the same position (34).

The trajectory of the DNA in the Nogales structure is blocked by the C-terminal domain of Ssl2; a high-frequency binding in the cleft of bacterial RNA polymerase, of Ssl2 was observed at the same position (34). Five intermediate complexes in transcription initiation.

Molecular basis of transcription initiation.

Arabidopsis thaliana TFIIJ.


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

Supplementary Materials and Methods

www.sciencemag.org/content/342/6159/1238724/suppl/DC1

Supplementary Materials and Methods

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

ARTICLE

RESEARCH

SCIENCE

www.sciencemag.org SCIENCE VOL 342 8 NOVEMBER 2013 1238724-7
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.1238724 Originally 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 TFIIA forming a bridge between them.