






Direct binding of TFE α opens DNA binding cleft of RNA polymerase

Sung-Hoon Jun ^{1,2,3,6,7}✉, Jaekyung Hyun ^{1,4,6}, Jeong Seok Cha ², Hoyoung Kim², Michael S. Bartlett⁵, Hyun-Soo Cho ^{2,7}✉ & Katsuhiko S. Murakami ^{3,7}✉

Opening of the DNA binding cleft of cellular RNA polymerase (RNAP) is necessary for transcription initiation but the underlying molecular mechanism is not known. Here, we report on the cryo-electron microscopy structures of the RNAP, RNAP-TFE α binary, and RNAP-TFE α -promoter DNA ternary complexes from archaea, *Thermococcus kodakarensis* (*Tko*). The structures reveal that TFE α bridges the RNAP clamp and stalk domains to open the DNA binding cleft. Positioning of promoter DNA into the cleft closes it while maintaining the TFE α interactions with the RNAP mobile modules. The structures and photo-crosslinking results also suggest that the conserved aromatic residue in the extended winged-helix domain of TFE α interacts with promoter DNA to stabilize the transcription bubble. This study provides a structural basis for the functions of TFE α and elucidates the mechanism by which the DNA binding cleft is opened during transcription initiation in the stalk-containing RNAPs, including archaeal and eukaryotic RNAPs.

¹Electron Microscopy Research Center, Korea Basic Science Institute, Chungcheongbukdo 28119, Republic of Korea. ²Department of Systems Biology, College of Life Science and Biotechnology, Yonsei University, Seoul 03722, Republic of Korea. ³Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA 16802, USA. ⁴Molecular Cryo-electron Microscopy Unit, Okinawa Institute of Science and Technology Graduate University, Okinawa 904-0495, Japan. ⁵Department of Biology, the Portland State University, Portland, OR 97207, USA. ⁶These authors contributed equally: Sung-Hoon Jun, Jaekyung Hyun. ⁷These authors jointly supervised this work: Sung-Hoon Jun, Hyun-Soo Cho, Katsuhiko S. Murakami. ✉email: jsh100@kbsi.re.kr; hscho8@yonsei.ac.kr; kum14@psu.edu

Transcription is the first step of gene expression, performed by cellular RNA polymerases (RNAPs) in the three life domains bacteria, archaea, and eukaryote. The structures of RNAP are highly conserved in archaea and eukaryotes, and consist of a “crab-claw” shape core with a protruding stalk domain^{1–3}.

For promoter-specific transcription, RNAP and general transcription factors (GTFs) are recruited to the promoter DNA to form the pre-initiation complex (PIC)⁴. The melting of DNA around the transcription start site (TSS) converts PIC to the open complex (OC), which is competent to initiate RNA synthesis. Archaeal RNAP requires three general transcription factors (GTFs)—TBP, TFB, and TFE α ⁵—whereas Pol II requires more complex GTFs including TFIIA, TFIIB, TFIID/TBP, TFIIE, TFIIF, TFIIH, and the Mediator complex^{4,6}. The orthologues of archaeal TBP, TFB, and TFE α are TBP, TFIIB, and TFIIE α -subunit, respectively, which indicates that the archaeal transcription system requires a subset of Pol II GTFs. Due to its simplicity, the archaeal transcription machinery is considered to be the evolutionary precursor and simplified form of the Pol II transcription apparatus that retains the conserved functions between two systems.

The DNA binding cleft of cellular RNAPs can adopt open or closed state by the conformational changes of the clamp domain of RNAP. Opening of the DNA binding cleft is required for promoter DNA loading into the cleft⁶. For Pol II, determination of PIC and OC structures was achieved in yeast and human^{7–9}. Yeast PIC structures show the closed cleft, whereas it shows the open cleft in human. Recent cryo-electron microscopy (cryo-EM) structure of yeast PIC reveals the closed clamp with a distorted promoter DNA in the closed cleft, which suggests that the transition of the clamp from the open conformation to the closed conformation induces DNA distortions to prime DNA melting¹⁰. In the OC structures of both human and yeast, the cleft is closed with the closed clamp conformation. In contrast, the biophysical study of archaeal transcription complexes using a single-molecule fluorescence energy transfer (smFRET) assay suggested a closed cleft PIC and open cleft OC¹¹. Together, the status of the cleft during the transcription initiation process for archaeal RNAP and Pol II are not established and the dynamic nature of the PIC structure^{7,10} presents a challenge to answering this question.

TFE α is a conserved GTF in archaeal-eukaryal transcription systems^{12,13} and is an essential factor in vivo¹⁴. It consists of an N-terminal extended winged-helix domain (eWHD) and a C-terminal zinc ribbon domain (ZRD), and the X-ray crystal structure of the eWHD was determined from *Sulfolobus solfataricus* (*Sso*)¹⁵. TFE α enhances transcription efficiency of RNAP in vitro, particularly for weak promoters under sub-optimal conditions^{16–18}, facilitates the formation of PIC and promoter DNA melting to form OC¹⁷, and also stabilizes OC¹⁹. In relation to the clamp conformation, the smFRET study suggested that TFE α allosterically aids in opening the clamp during OC formation¹¹. However, the molecular mechanisms of the functions of TFE α remain elusive due to lack of information about its structural conformation in the transcription complexes.

TFIIE is a heterodimer complex of TFIIE α/β in human and Tfa1/2 in yeast. The N-terminal region of TFIIE α or Tfa1 is essential and corresponds to archaeal TFE α ²⁰, while TFIIE β or Tfa2 is conserved only in eukaryotes and is functionally redundant^{6,13}. The interactions between Pol II and TFIIE in the PIC have been studied from the cryo-EM structures in human and yeast^{7–9} but the dynamic and transient nature of the PIC limits the resolution of the structures. Yeast PIC structures reveal the closed clamp, and there is no direct interaction between TFIIE and the clamp^{7,9}, whereas the human PIC structure is in the open clamp, enabling TFIIE interaction with the clamp⁸. Human PIC

structure suggests that TFIIE β interacts with the clamp⁸. Additionally, biochemical probes and crosslinking suggest that Tfa1 interacts with the clamp in PIC^{21,22}. Thus, the molecular details of interactions between TFIIE and Pol II in the PIC remain unresolved and the molecular mechanism of how TFIIE is involved in the clamp conformation change is unknown.

In this study, we determined the X-ray structure of TFE α and the cryo-EM structures of three forms of archaeal RNAP, including the apo-form, RNAP–TFE α binary complex, and RNAP–TFE α –DNA ternary complex. The structures reveal that direct binding of TFE α can shift the clamp conformation from a closed conformation to an open conformation. Our results show dynamic structural interactions among TFE α , RNAP, and promoter DNA, which provides structural information for the TFE α functions and transcription mechanism.

Results

X-ray crystal structure of TFE α . To determine the structure of TFE α , we first tried, unsuccessfully, to crystallize full-length TFE α from euryarchaea *Thermococcus kodakarensis* (*Tko*) and *Pyrococcus furiosus* (*Pfu*). After screening derivatives of TFE α , we found that the 13 kDa fragment *Pfu* TFE α (TFE α Δ ZRD, 1–110 amino acid residues) was crystallized. The structure was solved by a single wavelength anomalous dispersion (SAD) at 3.2 Å and the resolution of the structure was improved to 2.6 Å with a native crystal (Supplementary Table 1). The TFE α Δ ZRD structure consists of the N-terminal eWHD (residues 5–86) and a long α -helix (α 5) that connects the eWHD and the ZRD (Fig. 1a, b). The eWHD structure is similar with the X-ray crystal structures of the eWHD of *Sso* TFE α ¹⁵ and the eWHD of human TFIIE α ²³.

We find previously unrecognized features of the TFE α structure from the TFE α Δ ZRD structure. There were eight molecules in the asymmetric unit of the SAD structure and two molecules in the asymmetric unit of the native structure. Superposition of the globular part of the eWHD from the 10 crystallographically independent structures (average root-mean-square deviation in α -carbon positions are 0.369–0.589 Å over residues 8–62) reveal flexibilities in the wing and the α 5 linker (Fig. 1c). The conformation of the wing can change about 24° (~13 Å), and the α 5 can be bent at ~21° (~12 Å) for the eWHD (Fig. 1d). These results show a potential structural adaptability of TFE α during the transcription initiation process.

Cryo-EM structures of apo-RNAP and the binary complex of RNAP–TFE α . To investigate molecular details of the interactions among RNAP, TFE α , and promoter DNA during transcription initiation, we determined the cryo-EM structures of *Tko* RNAP in three distinct states in the transcription cycle (Fig. 2a, b). The three forms were the apo-form, the TFE α bound form (RNAP–TFE α binary complex), and the TFE α and DNA bound form (RNAP–TFE α –DNA ternary complex) at nominal resolutions of 3.9 Å, 4.0 Å, and 3.8 Å, respectively (Supplementary Table 2 and Supplementary Figs. 1–3).

We previously solved the X-ray crystal structure of *Tko* RNAP². To resolve accurate structural changes of *Tko* RNAP in solution through interactions with TFE α and promoter DNA, we determined the cryo-EM structure of *Tko* apo-RNAP. Three-dimensional classification of the apo-RNAP dataset showed that 82% particles formed complete apo-RNAP structures whereas 18% particles formed a stalkless RNAP structure (Supplementary Fig. 1c). The class of apo-RNAP comprising 56%, which we refer to as a major class of apo-RNAP, was further refined and the resulting map showed a well-resolved core and structural periphery. We modeled the apo-RNAP structure into the cryo-EM map (Supplementary Fig. 1f) and compared it with the

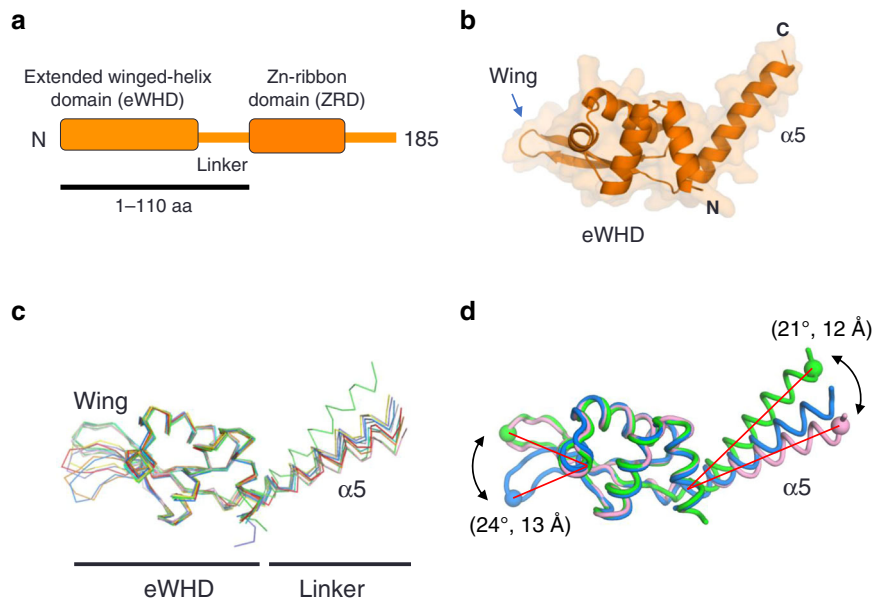


Fig. 1 X-ray crystal structure of *Pfu* TFE α Δ ZRD. **a Domain organization of the TFE α . The amino acid region used for crystallization is underlined. **b** A ribbon model of TFE α Δ ZRD is shown with transparent surface. **c** Ten crystallographically independent TFE α Δ ZRD structures are shown as α -carbon backbone. The structures are superimposed via eWHD with wing excepted, showing flexibility in the orientations of the wing and linker. **d** Flexibilities of the wing and linker of TFE α . Three representative conformations of crystallographically independent TFE α Δ ZRD show swing motions of the wing and linker of TFE α .**

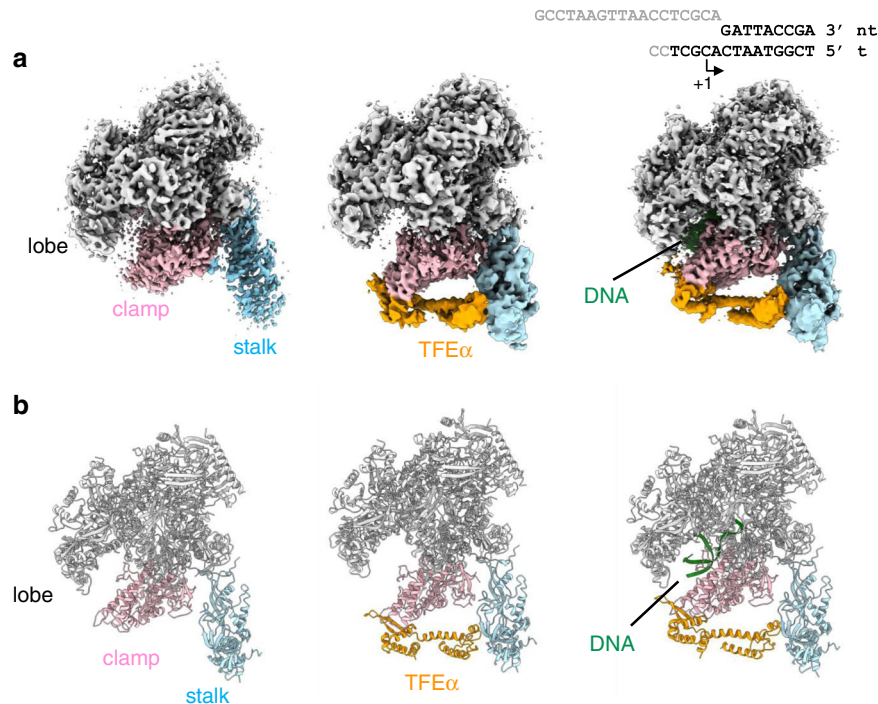
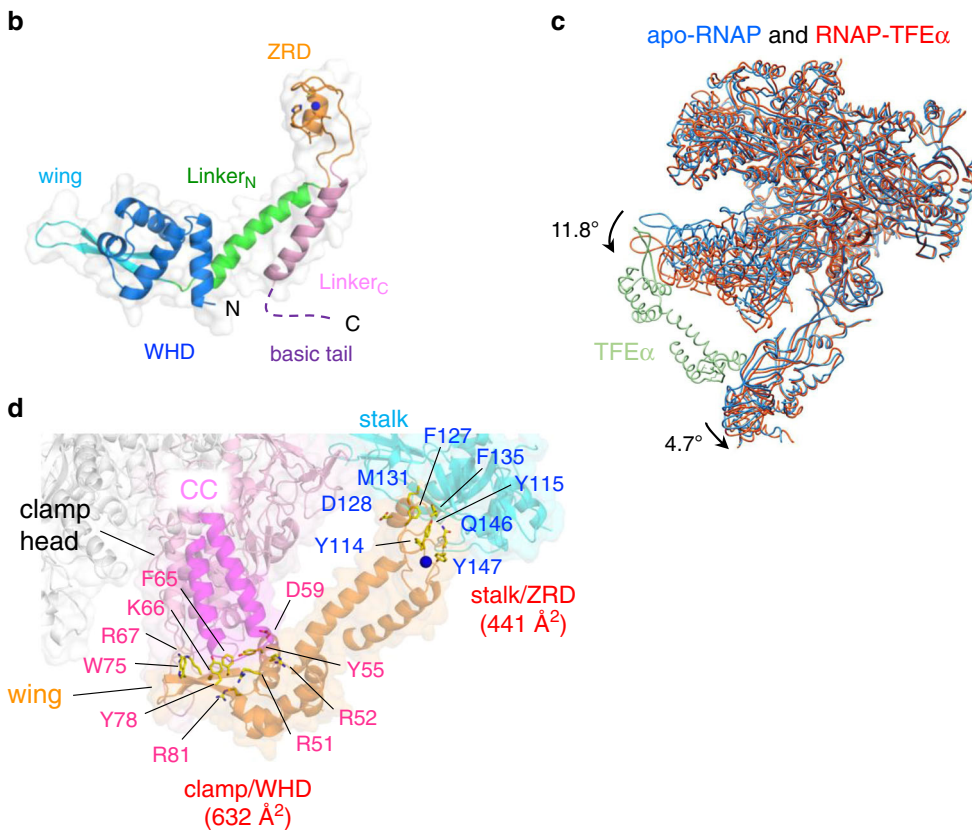
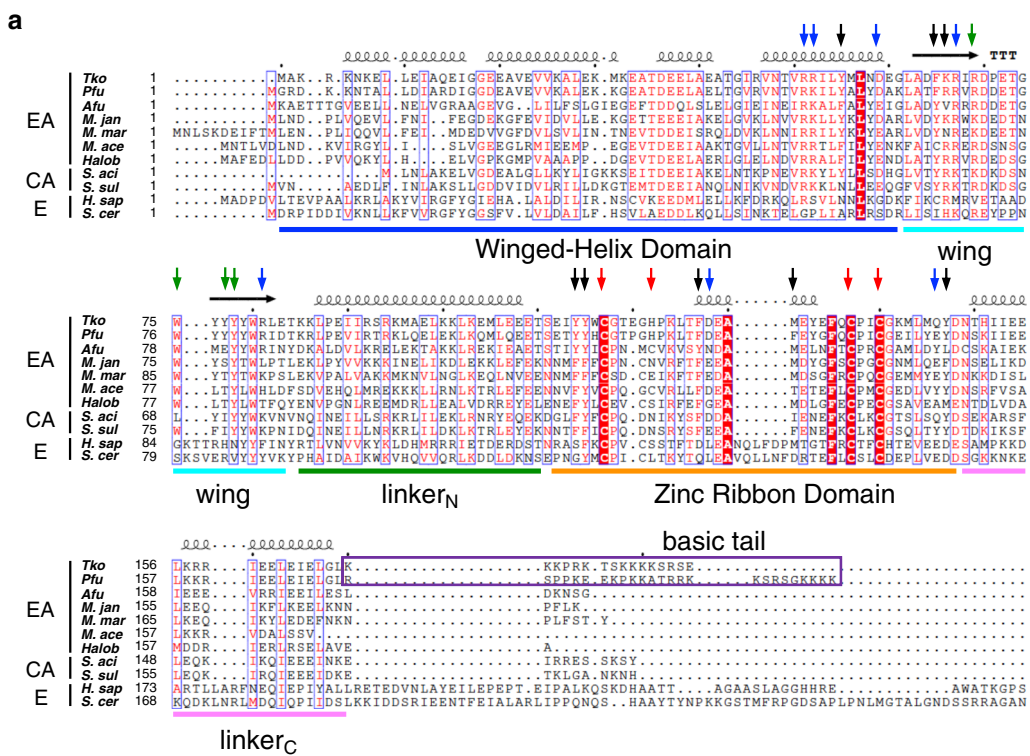


Fig. 2 Cryo-EM structures of three different forms of the *Tko* RNAP. **a Cryo-EM density maps of the apo-form (left), RNAP-TFE α binary complex (center), and RNAP-TFE α -DNA ternary complex (right). Domains of RNAP and DNA are indicated and variably colored. The DNA used for preparing the ternary complex is shown (right). DNA bases shown in gray are disordered in the structure. Note that the cryo-EM maps of RNAP-TFE α binary complex and RNAP-TFE α -DNA ternary complex are the composite maps produced from globally refined maps and focused refinements on the region containing TFE α , stalk and cleft domains. **b** Ribbon models of the apo-form (left), binary complex (center) and ternary complex (right) in the same color and orientation as in (a).**

previous X-ray crystal structures of archaeal RNAP^{1,2}. We discovered that the cryo-EM structure of apo-RNAP adopts a closed clamp conformation (Supplementary Fig. 4). Another 3D class covering 22%, referred as a minor class, was also refined and

the resulting structure was similar with the major class structure with 3.3° shift in clamp conformation to the open direction compared to the major class structure, indicating that there is a slight mobility in the clamp of apo-RNAP. Consistent with this,



the results of the smFRET experiment suggested that free apo-RNAP in solution is predominantly in a closed clamp conformational state¹¹.

To visualize the effect of TFE α binding on RNAP and molecular details of the RNAP and TFE α interaction, we determined the cryo-EM structure of the RNAP-TFE α binary

complex. The amino acid sequences of TFE α from *Pfu*, and *Tko* are 71% identical (Fig. 3a), which allowed us to use the X-ray structure of TFE α Δ ZRD determined in this study (Fig. 1b) as a guide to build the complete model of TFE α with the eWHD, ZRD, and two α helix bundles linking these domains (Figs. 2b and 3b). The overall shape of *Tko* TFE α structure resembles the

Fig. 3 RNAP and TFE α interaction. **a** Amino acid sequence alignment of earchaeal (EA) and crearchaeal (CA) TFE α and eukaryotic (E) TFIIIE α . The amino acid residues corresponding to the domains and motifs are indicated by bars and are the same color as in (A). Absolutely conserved residues are shown as white letters with red background, and highly conserved residues are indicated by red letters. Secondary structures of *Tko* TFE α are shown (α helix, coil; β strand, arrow). Amino acid residues involved in the RNAP interaction are indicated by arrows of the same color as in (B). *Tko*, *Thermococcus kodakarensis*; *Pfu*, *Pyrococcus furiosus*; *Afu*, *Archaeoglobus fulgidus*; *M.jan*, *Methanocaldococcus jannaschii*; *M.mar*, *Methanococcus maripaludis*; *M.ace*, *Methanosarcina acetivorans*; *Halob*, *Halobacterium*; *S.aci*, *Sulfolobus acidocaldarius*; *S.sol*, *Sulfolobus solfataricus*; *H.sap*, *Homo sapiens*; *S.cer*, *Saccharomyces cerevisiae*. **b** TFE α structures in the RNAP–TFE α binary complex are depicted as a ribbon model and transparent surface. Domains and motifs are indicated and colored as in (a). Disordered C-terminal tail is shown as a dashed line. **c** The structures of the apo-RNAP (blue) and the RNAP–TFE α binary (red) complex are superimposed. TFE α in the binary complex is in soft green and conformational changes of the clamp and stalk by TFE α binding are indicated. **d** Close-up view of the TFE α and RNAP interaction. The ribbon models of RNAP (clamp, pink; coiled coil (CC), magenta; stalk, cyan) and TFE α (orange) are shown with transparent surfaces. Amino acid residues involved in the RNAP and TFE α interaction are depicted as sticks and labeled (magenta and dark blue for residues interacting with the clamp and stalk domains, respectively). The contact surface areas between the eWHD and clamp and between the ZRD and stalk are indicated.

TFIIIE α within the human PIC structure⁸. The distinction between the two is that TFE α in the RNAP–TFE α binary complex interacts directly with the tip of the clamp whereas there is no direct interaction between TFIIIE α and the tip of clamp in the human PIC (Supplementary Fig. 5)⁸. For coordinating Zn atoms in the ZRD, *Tko* TFE α uses three Cys residues (C117, C137, and C140) together with a His residue (H122) instead of using four Cys residues in case of the eukaryotic TFIIIE α (Supplementary Fig. 5a). Archaeal TFE α contains a positively-charged tail (16 residues) at its C-terminus but its density is not traceable (Fig. 3a, b).

To reveal the precise effect of TFE α binding on RNAP, we compared the cryo-EM structures of the apo-RNAP and RNAP–TFE α binary complex. In the latter, the clamp swings away from the DNA binding cleft at 11.8° relative to the apo-form RNAP, and this movement is coupled with the stalk swinging (4.7°) away from the clamp (Fig. 3c and Supplementary Movie 1). The structure shows that TFE α connects two mobile modules of RNAP, the clamp and stalk. It demonstrates that direct binding of TFE α on RNAP shifts the clamp and the stalk conformation from a closed state to an open state.

The RNAP–TFE α binary complex structure shows the molecular details of the interplays between RNAP and TFE α to open the DNA binding cleft of RNAP. Two α helices between the eWHD and ZRD form an α -helix bundle linker, placing the ZRD over 50 Å away from the eWHD and allowing the eWHD and ZRD of TFE α to locate adjacent to the clamp and stalk domains of RNAP, respectively (Fig. 3d). The hydrophobic patch of the eWHD (residues Y55, F65, W75, and Y78) faces the tip of the coiled-coil (CC) of the clamp domain and the wing fits into the cavity between the CC and clamp head domain. Several hydrophilic residues of TFE α (R51, R52, D59, K66, R67, and R81) are also involved in the interaction with the RNAP clamp domain. The ZRD establishes its interaction with RpoE in the stalk domain of RNAP via hydrophobic surfaces (residues Y114, Y115, F127, M131, F135, and Y147). The D128 amino acid of TFE α also forms a salt bridge with K152 of RpoE. Consistent with the structure, deletion of the coiled-coil tip, amino acid substitutions at hydrophobic patch of RpoE (Y95E/F98E/L107E), or deletion of a short loop of RpoE (Δ S150–R157) disrupts TFE α binding to RNAP^{16,24}. The contact surface areas between the eWHD and clamp and between the ZRD and stalk are 632 and 441 Å², respectively. Amino acid residues involved in the TFE α –RNAP interaction are conserved in archaeal TFE α and only a few are conserved in eukaryotic TFIIIE α (Fig. 3a). This explains the much lower interaction affinity in eukaryotic Pol II counterparts.

The structure of RNAP–TFE α –DNA ternary complex: DNA loading to the RNAP cleft closes the clamp. To examine the conformation of archaeal RNAP with TFE α in the DNA bound

form, and how TFE α stabilizes the RNAP and DNA complex, we determined the cryo-EM structure of RNAP–TFE α –DNA ternary complex (Supplementary Table 2). We used a synthetic scaffold DNA in the form of a downstream fork junction promoter DNA containing the downstream double-stranded DNA and the partial transcription bubble (Fig. 2a), which functionally mimics open promoter DNA²⁵. Strong densities are observed in the ternary complex for double-stranded DNA within the DNA binding cleft and 5 bases of single-stranded template DNA (from –3 to +2) positioned near the active site of RNAP. However, density for the non-template DNA in the bubble is not traceable (Fig. 4a). The DNA position within the RNAP DNA binding cleft is the same as that of the cryo-EM structures of the yeast and human OC^{7,8}.

To identify the structural changes of RNAP and TFE α upon the binding of the transcription bubble-mimicking promoter DNA, we compared the cryo-EM structures of the RNAP–TFE α binary complex and RNAP–TFE α –DNA ternary complex and found that DNA binding to the DNA binding cleft of RNAP changes the conformations of the clamp and stalk back to the closed state (Fig. 4b). Although the clamp and stalk swing 17° (14 Å) and 7.3° (3.7 Å), respectively, TFE α still maintains the same interactions with these domains, as observed in the RNAP–TFE α binary complex. Relative orientations of the eWHD and ZRD in TFE α change in the binary and ternary complexes (Fig. 4c). The structural flexibility of TFE α , including the α 5 linker (Fig. 1c, d), allows it to bind the domains and function dynamically in both open and closed states.

Closing of the clamp establishes a new interaction between the TFE α wing and the lobe domain of RpoB, including the interaction of R69 side chain (TFE α) with Q285 side chain and G283 main chain at the lobe loop (Fig. 4d). This interaction encloses the DNA binding main cleft and provides a structural basis for TFE α stabilization of the OC, which is reminiscent of the mechanism for Spt4/5 stabilizing the elongation complex (Supplementary Fig. 6)²⁶.

Models of the PIC and OC. We modeled the archaeal PIC and OC (Fig. 5 and Supplementary Movie 2) using the cryo-EM structures of the RNAP–TFE α and RNAP–TFE α –DNA complexes determined in this study. For modeling DNA, TBP, and TFB, we used the cryo-EM structures of the yeast PIC containing a straight B-form DNA (PDB: 6GYK)¹⁰ and human OC (PDB: 5IYB) containing both the template and non-template DNA strands in the transcription bubble⁸ as templates. Cryo-EM structures of the *Tko* RNAP–TFE α (open clamp form) and RNAP–TFE α –DNA complexes (closed clamp form) were aligned with Pol II in the PIC and OC, respectively, using the catalytic core of RNAP containing the double-phi- β -barrel (DPBB) domains of the largest (RpoA1/Rpb1) and the second largest (RpoB/Rpb2) subunits.

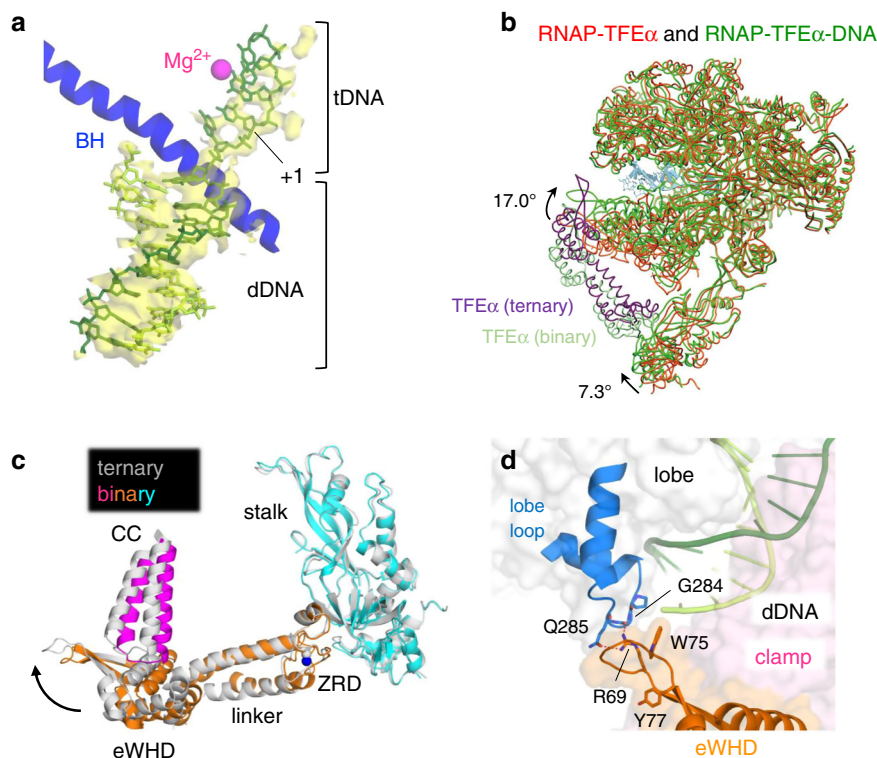


Fig. 4 DNA interaction with RNAP in the presence of TFE α . **a** Close-up view of the DNA binding channel of the RNAP-TFE α -DNA complex is shown with the cryo-EM density map (yellow and transparent) of DNA. DNA is shown as a stick model (template DNA, dark green; non-template DNA, light green). Template DNA strand (tDNA), downstream double-stranded DNA (dDNA) and transcription start site (+1) are indicated. Mg $^{2+}$ bound at the active site of RNAP and bridge helix (BH) are depicted. **b** The structures of the RNAP-TFE α binary (red) and the RNAP-TFE α -DNA ternary (green) complexes are superimposed. The conformational changes of the clamp and stalk by DNA positioning in the ternary complex are indicated. **c** Conformational change of TFE α upon RNAP binding DNA. The ZRD of TFE α was used as a reference to superimpose the RNAP and TFE α in the binary (color) and ternary complexes (light gray), revealing movement of the eWHD of TFE α together with the RNAP clamp indicated by an arrow. **d** Close-up view of the RNAP lobe and TFE α wing interaction. Amino acid residues involved in the RNAP lobe and TFE α wing interaction are depicted as stick models and indicated. Surface exposed aromatic residues (W75 and Y77) of TFE α involved in the DNA interaction are also indicated.

In the PIC model, the open clamp state of RNAP can accommodate double-stranded DNA between the TFE α wing and RpoB lobe loop (Fig. 5a), whereas clamp closure results in a steric clash between the TFE α wing and DNA. The lobe loop positions right below a major groove of DNA around $-3T/-2T$ position, whereas the TFE α wing does not establish any contact with DNA. Thus, the opening state facilitates the subsequent positioning of the double-stranded downstream DNA into the DNA binding cleft, which explains the transcription enhancing activity of TFE α .

In the OC model, the RNAP clamp can be closed without steric clash with single-stranded DNA strands in the transcription bubble (Fig. 5b). The TFE α wing is located adjacent to the non-template strand around the upstream edge of the transcription bubble, which is similar to the eukaryotic OC structures^{7,8}. We observe that $\beta 2$ of the wing is on the side of the non-template DNA and the aromatic side chain of W75 is positioned to establish the stacking interaction with the base of the non-template DNA at -9 position where the DNA bubble starts. This is reminiscent of the interaction between the bacterial RNAP σ factor region 2.3 and the non-template DNA in the open promoter complex, RPO²⁷. This suggests a molecular mechanism of how TFE α stabilizes the transcription bubble during the transcription initiation by the interaction with the upstream DNA of OC¹⁹.

Crosslinking experiments in the archaeal PIC and OC. To investigate how TFE α interacts with promoter DNA and the non-template DNA strand observed in the OC model, we used a

highly specific UV-inducible crosslinking system based on the site-specific incorporation of unnatural amino acid para-benzyoyl-phenylalanine (Bpa)²⁸ (Fig. 6). We prepared a series of Bpa substituted *Pfu* TFE α variants, including Bpa substitutions at I17 and N49 in the globular part of eWHD and Y56, F66, R70, W76, Y78, and Y79 in the wing (Fig. 6a). The Bpa reacts preferentially with unreactive C-H bonds when exposed to long wave UV light, with a reactive spherical radius of 3.1 Å²⁹, and these covalent cross-links can provide very specific proximity information for interacting macromolecular surfaces.

First, to establish whether the TFE α Bpa substitutions affected transcription activation by TFE α , we examined each TFE α variant for *gdh* promoter-dependent transcription in transcription complexes containing TBP and TFB2 (Fig. 6c and Supplementary Fig. 7). TFB2 was used since TFE α activation is greater with TFB2 than with TFB1³⁰. A modest activation was observed for wild type TFE α and all Bpa variants except for Y79Bpa, which lacked any activation, possibly because of interference by the bulkier Bpa R-group. This indicated that the transcription activation by TFE α was not affected by Bpa substitutions in most variants.

Next, to monitor the site-specific contact between TFE α and DNA, we prepared three sets of *gdh* promoter DNA containing radiolabel at $-9/-8/-7$ positions of non-template DNA ($-9NT$), at -3 position of non-template DNA ($-3NT$), and at -4 position of template DNA ($-4T$) (Fig. 6b). The $-9NT$ and $-3NT/-4T$ promoter positions are located at the upstream edge and at the middle of the DNA bubble in the open complex of the eukaryotic Pol II transcription system^{7,8}. Previous DNA to protein cross-links

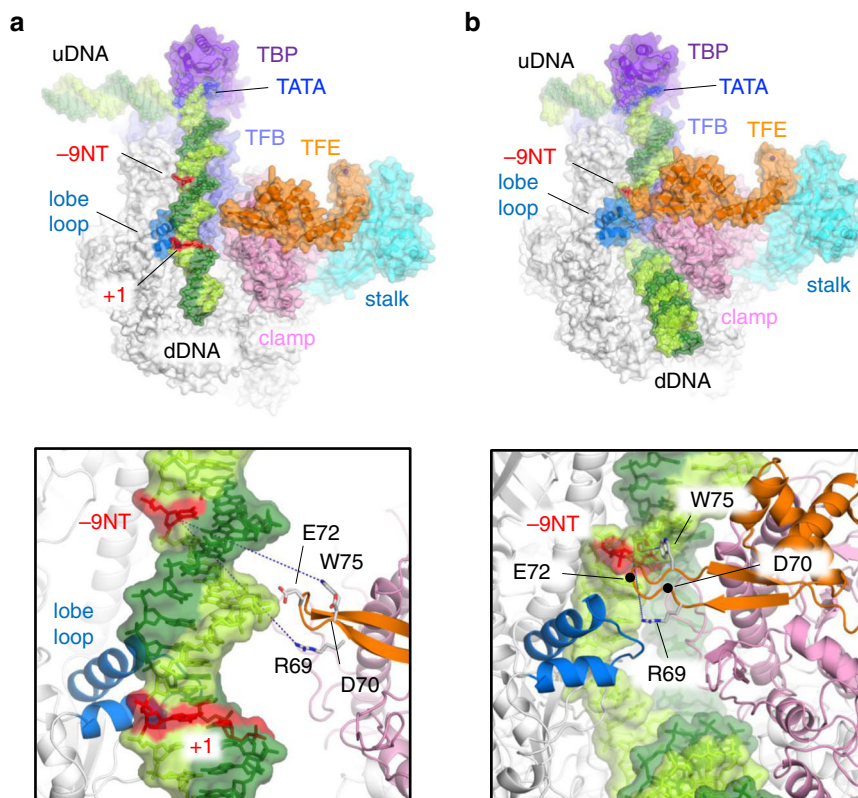


Fig. 5 Structural models of archaeal PIC and OC models. Overall views of the PIC (**a**) and the OC (**b**). RNAP (white with pink, cyan, and light blue), TBP (purple), TFB (light purple) and TFE α (orange) are depicted as ribbon models with transparent surface. DNA is depicted as stick model with transparent surface (template DNA, dark green; non-template DNA, light green; TATA box, blue). Positions of -9 base of non-template DNA (-9 NT) and transcription start site ($+1$) are indicated. Close-up views of DNA from -9 to $+1$ positions in the PIC and RPo are shown in boxes. Amino acid residues involved in the DNA interaction are depicted as stick models and labeled. TFB is removed to clarify the view.

indicated that TFE α in preinitiation complexes cross-links strongly to -11 and -9 of the NT strand, but not to the T strand^{19,31}. We observed crosslinking signals for the DNA radiolabelled at -9 NT position with TFE α variants with Bpa at R70, W76, Y78, and Y79 (Fig. 6d and Supplementary Fig. 8). The strongest signal was observed in W76Bpa, located at the tip of wing, indicating that W76 is positioned proximally to the upstream edge of transcription bubble. This result is consistent with the observation in the OC model (Fig. 5b) that the corresponding residue W75 is located to interact with the non-template DNA at the upstream edge of transcription bubble. In the OC model, when the clamp is open, the distance between -9 NT base and W75 of TFE α is 21 Å, which is too far for crosslinking, whereas the distance becomes 10 Å when the clamp is closed. Thus, the crosslinking result also indicates that the RNAP clamp is closed in the OC.

Discussion

Here, we report the cryo-EM structures of apo-RNAP, RNAP-TFE α binary complex, and RNAP-TFE α -promoter DNA ternary complex that visualize dynamic structural changes of TFE α and archaeal RNAP, especially in the clamp and stalk, during transcription initiation (Supplementary Movies 1). Stable binding of TFE α to *Tko* RNAP without other GTFs and promoter DNA³² allowed us to demonstrate that the direct binding of TFE α opens the DNA binding cleft of RNAP. We suggest a simplified transcription initiation model (Supplementary Movie 2) in archaea based on the model structures of PIC and OC in this study (Fig. 5). In this model, the clamp of apo-RNAP is closed conformation with a slight mobility. In the PIC, the binding of

TFE α on the clamp and stalk shifts the conformations of the mobile domains of RNAP to an open state (Fig. 5a) and then, the double-stranded downstream promoter DNA is placed into the DNA binding cleft. The positioning of the downstream DNA into the DNA binding cleft induces the clamp closing by strong ionic interactions between positively-charged residues covering the inside of the DNA binding cleft and a negative-charged DNA backbone. During the clamp closure, DNA distortion may be induced as observed from the yeast Pol II system¹⁰, and we suggest that the wing of TFE α may be inserted into the DNA minor groove around -6 position where the transcription bubble is formed (Supplementary Movie 2). This also explains how TFE α facilitates the unwinding of the DNA. In the OC, the clamp is in a closed state and the wing of TFE α is located to stabilize OC structure by its interactions with the opposite region of the cleft (Supplementary Fig. 6) and the non-template strand DNA at the upstream edge of the transcription bubble (Fig. 5b).

This model shows how the DNA binding cleft of stalk-containing RNAPs can be opened in PIC and provides the structural basis of TFE α functions to facilitate transcription initiation. A previous study using smFRET assay suggested TFE α helps clamp opening allosterically during OC formation¹¹. However, our results show that the direct binding of TFE α opens the clamp conformation. In the cryo-EM structures of eukaryotic Pol II PIC, the interaction between Pol II and TFIIE was not resolved in relation with the conformation of the clamp and stalk because of the dynamic and transient nature of PIC structure as well as the complex composition⁷⁻¹⁰ (Supplementary Fig. 9). Given that the essential N-terminal part of TFIIE α in humans (Tfal in yeast) corresponds to TFE α ²⁰ and their structures are

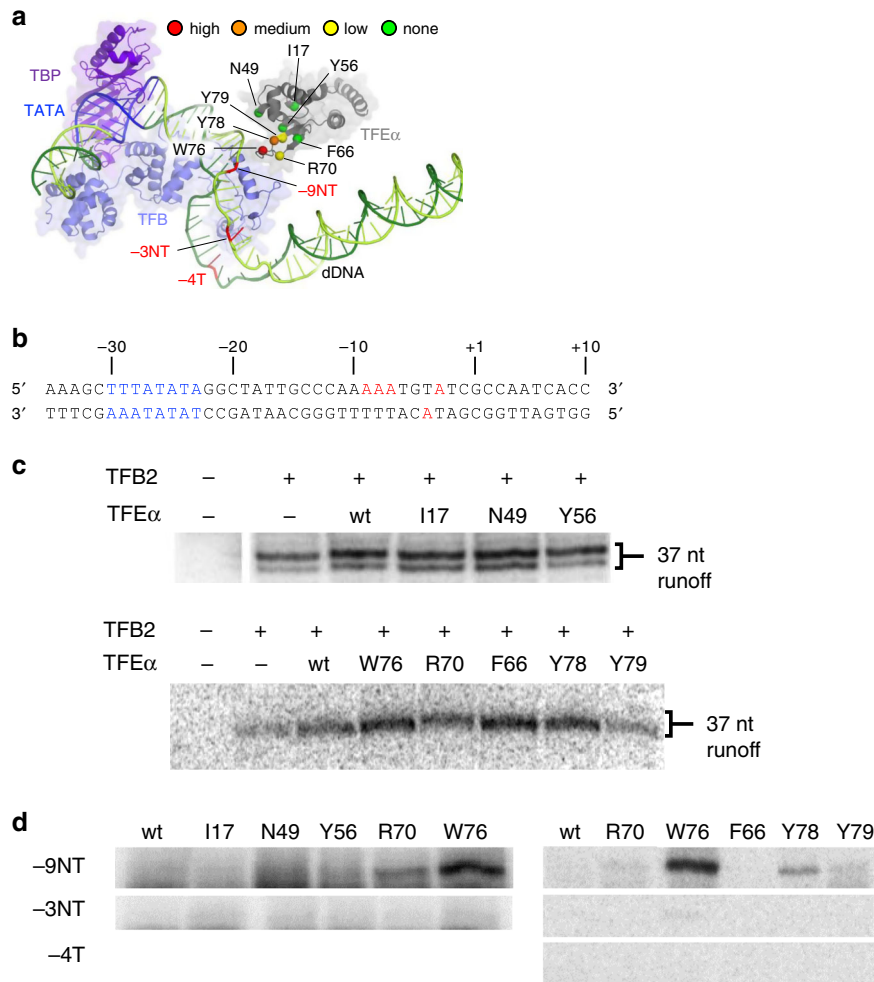


Fig. 6 TFE α and DNA interaction investigated by photo-crosslinking. **a** Structures of TBP, TFB, TFE α , and DNA in the OC model (Fig. 5b). Bpa substituted amino acids of TFE α (*Pfu* numbering) are indicated. Colors indicate cross-linking efficiency to DNA position at -9NT. Positions at -9NT, -3NT, and -4T are indicated. **b** DNA sequence of the *gdh* promoter used for photo-crosslinking. TATA box and radiolabelled positions are highlighted by blue and red, respectively. **c** Effects of Bpa substitution on transcription. Transcription of the *gdh* promoter was done in the absence (--) or presence of TFB2 and wild type and Bpa-substituted TFE α variants. Two sets of experiments are shown. In the upper set, run-off transcription gave rise to a primary transcript and a slightly shorter transcript, likely due to heterogeneity in the downstream end of the *gdh* promoter PCR product used. Full scan of the figures is in Supplementary Fig. 7. Each TFE α variant was tested at least two times, with representative results shown. **d** Crosslinking of TFE α Bpa variants to the *gdh* promoter. Preinitiation complexes were formed using *gdh* promoter variants radiolabelled on the NT strand (-9, -8, -7 or -3) or the T strand (-4), and cross-linking triggered by UV light. Cross-linked DNA was processed by nuclease and TFE α variants that retained radiolabel (indicating proximity between Bpa and radiolabel) were identified by SDS-PAGE. The region of the gels spanning ~25 to 30 kDa is shown. Full scan of the figures is in Supplementary Fig. 8. Each TFE α variant was tested with each radiolabelled *gdh* promoter variant at least three times, with representative results shown.

conserved (Supplementary Fig. 5), our results suggest that the clamp opening activity of TFE α in the PIC by the direct interaction with the tip of the clamp is conserved in TFIIE α (Tfa1 in yeast) as well. This is consistent with other studies^{8,33} where the intermediate PIC structure having an open clamp conformation has been observed in humans, and has been reported as a requirement for the priming of DNA melting¹⁰. Additionally, our results are concurrent with previous findings that the cross-linking assay of yeast PIC showing eWHD of Tfa1 is located close to the clamp in the PIC²².

Our results also suggest how the binding of TFE α to RNAP can shift the conformation of the clamp and stalk from a closed state to an open state. The cryo-EM reconstruction of the RNAP-TFE α binary complex in the high threshold of the map shows that the density between the ZRD and the stalk is rigid, whereas the density of the eWHD and clamp is weak (Supplementary Fig. 10), which suggests a modest structural heterogeneity in the eWHD and clamp. The asymmetric interaction of TFE α on the stalk and

clamp shows the mode of TFE α binding on RNAP and suggests a model of how TFE α binding on the clamp and stalk domains shifts their conformations from a closed state to an open state. In this model, the interaction between the ZRD and the stalk tethers TFE α on the RNAP. Subsequently, the eWHD of TFE α interacts with the coiled-coil tip of the clamp, which physically connects the center region of the stalk and the tip of the clamp. The bridging of the two domains of RNAP by TFE α mechanically allows the conformational change of the clamp from a closed state to an open state via a leverage effect. There is a slight structural heterogeneity in the clamp and eWHD due to the flexibility of the domains. The conformational change in the clamp produces a coupled conformational change in the stalk, as observed previously in other RNAP structures^{2,34}.

In eukaryotic Pol II, a clash of synthesized RNA with TFIIB causes promoter escape³⁵ and the subsequent initiation-elongation transition alters the path of upstream DNA without significant conformational change in Pol II^{7,35,36}.

The conformation of RNAP in the RNAP–TFE α –DNA ternary complex is in the closed clamp state (Fig. 4b) as archaeal RNAP in the RNAP–Spt4/5 complex structure²⁶ as well as Pol II in the OC structure^{7–9} and the transcription elongation complex (EC) structure³⁶. TFE α can bind to RNAP stably in the RNAP–TFE α –DNA ternary complex structure (Figs. 2 and 4), which supports the notion that TFE α is a part of early EC¹⁹ and Spt4/5 displaces TFE α in the EC to enhance transcription processivity of the EC¹⁶.

smFRET study in solution characterized the conformational positioning of the archaeal RNA polymerase clamp during the transcription cycle¹¹. The structural information for the clamp conformation in this report is consistent with the smFRET work (Figs. 3c and 5a) but there is a difference after DNA melting and the OC formation (Figs. 4b and 5b). Recent reports showed that the OC formation is a multi-step process and clamp conformation changes are crucial event during DNA melting and the OC formation^{37,38}. The structures of complete archaeal transcription complexes including GTFs such as TBP and TFB along with further biophysical studies would give refined models of transcription initiation mechanism in archaea.

The simpler bacterial system suggested detailed models for promoter DNA melting and subsequent arrangement of the transcription bubble^{37–39}. Biophysical study assigned bacterial RNAP clamp conformation at each steps in transcription initiation and elongation³⁹ and real-time fluorescent study monitored that a transient clamp closure is required to nucleate DNA melting³⁷. Universal structural features of RNAP within the DNA binding cleft play key roles in the late steps of promoter DNA melting³⁸ and clamp opening is required for the template DNA strand to enter into the active site of RNAP^{37,38}. Our results support the universally conserved mechanisms of transcription initiation in all three domains of life by showing that direct binding of TFE α , a conserved GTF in archaeal-eukaryal RNAP systems, can open the clamp of the stalk-containing RNAPs. In *Thermus* and *Mycobacterium*, interestingly, an additional GTF CarD stabilizes the OC at the upstream edge of the transcription bubble^{40,41} as TFE α does in archaea, which also suggests some domain- or lineage-specific transcription factors incorporated during evolution to carry out common functions.

Methods

TFE α cloning, expression, and purification. The DNA fragment encoding the N-terminal 110 amino acids of the *Pfu* TFE α (TFE α Δ ZRD) was cloned into a pSUMO vector (LifeSensors) with BsaI/XhoI restriction sites (Supplementary Table 3). The plasmid was transformed into BL21(DE3) CodonPlus cells (Stratagene), and the transformed cells were grown in 2 \times YT media at 37 °C until OD₆₀₀ reached 0.6. Protein expression was induced by adding 0.2 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) for 20 h at 20 °C. Harvested cells were suspended in buffer A (10 mM Tris-HCl (pH 8.0), 500 mM NaCl, 10 mM imidazole, and 5 mM β -mercaptoethanol) supplemented with 1 mM phenylmethylsulfonyl fluoride and lysed by sonication. After centrifugation, the supernatant was loaded onto a 5 ml Ni-NTA column (Qiagen) equilibrated with buffer A. After washing with 20 column volumes of buffer A containing 20 mM imidazole, proteins were eluted with buffer A containing 200 mM imidazole. The N-terminal (His)₆-SUMO tag was removed by the Ulp1 protease (LifeSensors) (1:200 of Ulp1:TFE α Δ C) for 1 hour at 20 °C. After dialysis against buffer A without imidazole, the TFE α Δ C was collected as the flow-through of a second Ni-NTA column. The protein was purified by heat treatment (70 °C for 30 min) and further purified by Superdex-200 size-exclusion column (GE Healthcare) equilibrated with buffer B (10 mM Tris-HCl (pH 8), 300 mM NaCl, 5% glycerol, and 5 mM dithiothreitol (DTT)).

Selenomethionyl (SeMet) *Pfu* TFE α Δ ZRD was prepared for the Single-wavelength anomalous diffraction (SAD) method. In addition to two original methionine residues at 1 and 106 positions, three methionines were introduced at residues Ile13, Ile54, and Phe66 positions using QuikChange site-directed mutagenesis (Stratagene). SeMet substituted protein was generated by growth in M9 minimal media supplemented with SeMet via the inhibition of methionine biosynthesis⁴², and purified as the native protein.

Crystallization and X-ray data collection. For crystallization trials, native and SeMet substituted *Pfu* TFE α Δ ZRD were concentrated to 6 mg/ml. Crystals were

grown within 2 days by hanging-drop vapor diffusion method at 21 °C with the crystallization solution containing 10% PEG 8000, 0.2 M MgCl₂, and 0.1 M Tris-HCl (pH 7). Crystals were cryo-protected by adding ethylene glycol to the reservoir solution to a final concentration of 20% prior to plunge-freezing in liquid nitrogen. Diffraction data were collected at the A1 beamline of the Cornell University High Energy Synchrotron Source (CHESS, Ithaca, NY) at 100 K and data were processed by HKL2000⁴³. Native crystals belonged to P4₁2₁2 and contained two molecules in an asymmetric unit, whereas SeMet labeled crystals belonged to P2₁2₁2₁ and contained eight molecules in an asymmetric unit.

X-ray crystal structure determination and refinement. With the anomalous signal from SeMet, 32 selenium sites in the asymmetric unit (4 sites in each of the 8 molecules in an asymmetric unit) were located and the experimental phase (figure of merit: 0.390) was calculated using Automated structure solution (AutoSol) in PHENIX⁴⁴. Density modification by Automated model building (AutoBuild) in PHENIX yielded an excellent map and most of the model was built automatically. Manual model building was done using Coot⁴⁵ nm, and the structure was refined using PHENIX. The structure of native *Pfu* TFE α Δ ZRD was determined by the molecular replacement using Automated molecular replacement (Phaser-MR) in PHENIX followed by the structure refinement. Final coordinates and structure factors have been deposited to the Protein Data Bank (PDB) with the accession codes listed in the Supplementary Movie 1.

Preparations of RNAP, RNAP–TFE α and RNAP–TFE α –DNA complexes. *Tko* RNAP was purified from Δ *rpo4* strain KUWLF16 and then the 11-subunit apo-*Tko* RNAP was reconstituted in vitro by supplementing recombinant Rpo4 and Rpo7². About 50 g of cells were suspended in 200 ml lysis buffer (10 mM Tris-HCl (pH 8.0), 500 mM KCl, 10% glycerol, 10 mM imidazole, 10 μ M ZnCl₂, 5 mM 2-mercaptoethanol, 0.3 μ M leupeptin, 1 μ M pepstatin, 1.5 mM benzamide hydrochloride and 0.5 mM phenylmethyl sulphonyl fluoride) and lysed by sonication. The supernatant was loaded to Ni-NTA affinity columns (Qiagen) equilibrated with the lysis buffer and washed with the same buffer containing 20 mM imidazole. Proteins were eluted with the lysis buffer containing 200 mM imidazole and precipitated by ammonium sulfate (final 80% saturation). The pellet was suspended in TGED buffer (20 mM Tris-HCl (pH 8.0), 10% glycerol, 0.5 μ M EDTA and 5 mM DTT) and RNAP was further purified by binding and elution from HiTrap Q HP (GE Healthcare). To reconstitute RNAP containing all subunits, RNAP pools from HiTrap Q were mixed with recombinant Rpo4/Rpo7 and Rpo4 at a ratio of 1:4:1 (RNAP:Rpo4/Rpo7:Rpo4) for 1 h at 20 °C and were further purified by successive passage and elution from HiTrap Heparin and Superdex200 columns (GE healthcare). The gene encoding *Tko* TFE α was cloned into a pSUMO vector (LifeSensors) with BsmBI/BamHI restriction sites (Supplementary Table 3) and *Tko* TFE α was expressed in BL21(DE3) CodonPlus cells (Stratagene) using a pSUMO vector (LifeSensors) in 2 \times YT media at 37 °C. The temperature was shifted to 18 °C when OD₆₀₀ of culture reached 0.6 and protein expression was induced by 0.2 mM IPTG for 20 h. After harvesting cells, TFE α was purified as TFE α Δ ZRD.

The binary complex of RNAP and TFE α was formed by mixing 3 nmole RNAP with 5-fold molar excess of TFE α at room temperature for 1 h. The sample was applied to a Superose 6 10/300 GL column (GE Healthcare) equilibrated with 10 mM Tris-HCl (pH 8.0), 400 mM NaCl, 5 mM DTT, 1% glycerol, 0.1 mM EDTA and collected in 300 μ l fractions. The fraction corresponding to the peak of the complex, in about 0.3 mg/ml, was used directly in the subsequent EM analysis.

For the preparation of RNAP–TFE α -promoter DNA ternary complex, DNA scaffold based on the *Tko glutamate dehydrogenase* (*gdh*) gene promoter (template strand: from –4 to +12, 5'-TCGGTAATCACGCTCC-3', non-template strand: from –15 to +12, 5'-GCCTAAGTTAACCTCGCAGATTACCGA-3') was used. Duplex DNA was generated by mixing equimolar single-stranded DNA oligonucleotides in H₂O and heating to 95 °C for 5 min. The sample was cooled to 10 °C at a rate of 1 °C per min.

The ternary complex of RNAP, TFE α , and DNA was prepared by mixing 3 nmol binary complex of RNAP–TFE α with 3-fold molar excess of the annealed DNA at room temperature for 30 min. The ternary complex was purified with a Superose 6 10/300 GL column (GE Healthcare) equilibrated with 10 mM Tris-HCl (pH 8.0), 400 mM NaCl, 5 mM DTT, 1% glycerol, 0.1 mM EDTA. A fraction of the peak of the complex in about 0.3 mg/ml was used in the EM analysis.

Cryo-electron microscopy data acquisition. Samples of 4 μ l of purified apo-RNAP, RNAP–TFE α , and RNAP–TFE α –DNA of ~0.3 mg/ml were loaded onto freshly glow-discharged holey carbon grids (Quantifoil R1.2/1.3 Cu200, Quantifoil Micro Tools GmbH, Germany) followed by plunge freezing using a Vitrobot Mark IV (Thermo Fisher Scientific Inc., USA) at 90% relative humidity and 4 °C. For the RNAP–TFE α –DNA sample, from which a strong orientation preference was observed, additional data acquisition was performed using the sample loaded onto a holey carbon grid coated with a graphene oxide film. Image data collection was performed using a Titan Krios transmission electron microscope (Thermo Fisher Scientific Inc., USA) operating at 300 kV acceleration voltage, and equipped with image spherical aberration (Cs) corrector (CEOS GmbH, Germany) and a Falcon II direct electron detector. Imaging was performed in nano-probe mode, spot size 3, 70 μ m C2 aperture and 100 μ m objective lens aperture. Cs corrector was tuned just

before data acquisition at normal magnification of $\times 125,000$ in order to minimize 2-fold and 3-fold astigmatism, axial coma, and spherical aberration. Automated data collection was performed using EPU software (Thermo Fisher Scientific Inc., USA) at a nominal magnification of $\times 47,000$, which corresponds to a pixel size of 1.4 Å at the specimen scale, and to a defocus range between $-1.5 \mu\text{m}$ and $-3.5 \mu\text{m}$. Thirty frames per movie data were collected at a dose rate of $\sim 35 \text{ e}^-/\text{Å}^2/\text{s}$ over a 1.8 s exposure time. A total of 1386 movies were collected for apo-RNAP and 2698 movies for RNAP-TFE α , respectively. For RNAP-TFE α -DNA complex, 1,737 movies and 2,106 movies were collected from the holey grid with and without additional graphene oxide support film, respectively.

Cryo-electron microscopy image processing. MotionCor2⁴⁶ was used for beam-induced movement correction and dose-weighting from the movie data. The motion-corrected micrographs were subjected to contrast transfer function estimation using Gctf⁴⁷. Relion 2.1⁴⁸ was employed for subsequent image processing procedures, unless otherwise stated. Typically, 10,000–20,000 particles were manually picked from micrographs with high underfocus to aid visibility of the particles. The particles were extracted into 176×176 pixel boxes followed by 2D class averaging. Class averages with pronounced structural features were used as a template for automated particle picking. The box size and particle-picking regime were commonly applied to all datasets.

For apo-RNAP, 514,007 particles were extracted from particle auto-picking. Two rounds of 2D class averaging were performed from which particles that belonged to class averages with poor structural features were eliminated after each run, leaving 350,067 particles in the dataset (Supplementary Fig. 1b). The X-ray structure of *Tko* RNAP (PDB: 4QIW) was converted to an EM map and low pass filtered to 60 Å, using the *e2pdb2mrc.py* routine in the EMAN2 software package⁴⁹. The volume was used as a reference for the subsequent 3D reconstruction process, but due to the strong preferential orientation of particles, the initial attempt of 3D reconstruction resulted in a poorly resolved map. To alleviate the effect of the preferred orientation, particle images that belonged to strongly preferred orientations were discarded based on the 2D class averages. The resulting 247,663 particles were used to reconstruct a consensus map without any preceding 3D classification. Then, the particles were classified into four 3D classes without performing any image alignment. The 139,242 particles that belonged in the 3D class that exhibited the best structural detail were selected for subsequent per-particle motion correction and B-factor weighing via particle polishing (Supplementary Fig. 1c). Finally, the corrected particle images were refined and post-processed to yield a reconstruction at 3.91 Å resolution based on a 0.143 Fourier shell correlation (FSC) cut-off criterion (Supplementary Fig. 1e)⁵⁰.

For RNAP-TFE α , a total of 1,226,339 particles were auto-picked. Then 953,770 particles belonging to the best 2D class averages were selected after two rounds of class averaging (Supplementary Fig. 2b). Using low pass filtered X-ray structure of *Tko* RNAP as an initial model, the dataset was subjected to 3D classification. Out of six resultant 3D classes, two classes exhibited the best structural details and unambiguously discernible densities for TFE α and stalk domain, and 409,653 particles belonging to those classes were selected for subsequent processing. The particles were re-classified in 3D without alignment in order to sort out particle images of intact complex and with best structural details. From eight 3D classes, 252,508 particles that belonged to the best two 3D classes chosen for final 3D reconstruction (Supplementary Fig. 2c). Particle polishing, refinement and post-processing resulted in a 3D reconstruction at 4.04 Å resolution based on a 0.143 FSC cut-off criterion (Supplementary Fig. 2e). A binary mask that embraced the density corresponding to stalk and clamp domains of RNAP and TFE α was created from a preliminary atomic model. Focused refinement on this region was performed by continuing the last iteration of the 3D refinement run of the entire structure with the applied local mask. The overall resolution of the resulting map from focused refinement was 4.25 Å with a slight improvement in local resolution of the masked region (Supplementary Fig. 2e, f).

For RNAP-TFE α -DNA, a total of 1,388,666 particles were auto-picked from the combined micrograph datasets obtained from the vitrified grids with and without graphene oxide film. After two rounds of 2D class averaging, 911,020 particles were selected (Supplementary Fig. 3a) and subjected to 3D classification into four classes. A total of 669,840 particles belonging to the 3D classes with pronounced structural details and intact domains, TFE α and DNA, were used to produce a consensus 3D reconstruction. The dataset was further classified into eight 3D classes without image alignment. After careful inspection of the 3D classes, two classes with intact TFE α , stalk domain, DNA, and good structural details were chosen, from which 312,092 particles were extracted (Supplementary Fig. 3b). Among the particles included for the final reconstruction, 121,727 and 190,365 particles were from the micrographs obtained from grids with and without graphene oxide film, respectively. After particle polishing, refinement, and post-processing, a final map at 3.79 Å resolution was produced (Supplementary Fig. 3d). A binary mask that embraces TFE α , DNA as well as stalk and clamp domains was applied for focused refinement that continued from the last iteration of the 3D refinement of the entire complex. The overall resolution of the resulting map from focused refinement was 3.85 Å with a slight improvement in local resolution of the masked region (Supplementary Fig. 3d, e).

Local resolution estimation of resulting cryo-EM map was performed using MonoRes by submitting half maps for the final reconstruction to the Scipion web

tools⁵¹. FSC between the unfiltered half maps was calculated using Relion 2.1. In case of RNAP-TFE α and RNAP-TFE α -DNA, the composite maps were produced by combining the final maps with the maps resulting from focused refinements using *combine_focused_maps* tool in PHENIX. FSC between the composite map and the refined atomic model was calculated using *phenix.mtriage* routine in PHENIX (Supplementary Figs. 2f and 3e)⁵². EM map display and figure preparation was performed using USCF Chimera⁵³.

Modeling of cryo-EM structures. We built molecular models of the apo-RNAP, RNAP-TFE α binary complex, and RNAP-TFE α -promoter DNA ternary complex into the cryo-EM maps using an X-ray crystal structure of apo *Tko* RNAP (PDB: 4QIW)². The X-ray crystal structure of *Pfu* TFE α Δ ZRD and that of human TFIIE (PDB: 5GPY)²³ were used as references. The *Tko* RNAP crystal structure was fitted to the map as a rigid body, and then subunits and the mobile clamp domain were fitted individually. We modeled and extended TFE α and DNA structures manually using Coot⁴⁵ and further real-space refined the structures against the B-factor sharpened cryo-EM maps from *relion_postprocess* with PHENIX suite⁴⁴ and validated using MolProbity⁵⁴ (Supplementary Table 2).

In vitro transcription and cross-linking. Mutant TFE α variants containing site-specific Bpa substitutions were made using the pSUP system^{28,55}. Amber codons were created in an N-terminal 6 \times -His tagged TFE α gene at desired amino acid positions, by primer extension/DpnI mutagenesis. The mutated TFE α variants were expressed from inducible plasmids alongside pSUP-BpaRS-6TRN, in the presence of *p*-benzoyl phenylalanine. The TFE α variants were then purified by nickel chromatography. Transcription reactions were conducted using purified recombinant TBP, TFB2, and TFE α and native *Pfu* RNAP (containing a 6 \times His tag on subunit D, purified by nickel chromatography⁵⁶). Proteins were assembled at the *gdh* promoter for 30 min at 65 °C, with RNAP at 10 nM, TBP at 60 nM, TFB2 at 120 nM, and TFE α at 200 nM³⁰. Ribonucleotide triphosphates (500 μM GTP, CTP, and ATP, 10 μM [α -³²P]UTP at $\sim 40 \text{ Ci mmol}^{-1}$) were added to start transcription, reactions were incubated for an additional 20 min at 65 °C, and then stopped by adding EDTA to 17 mM, along with a small amount of radiolabelled DNA marker (100 bp). Samples were phenol extracted and ethanol precipitated, and transcripts visualized on 14% polyacrylamide gels. See Supplementary Fig. 7 for full scans of the gels.

TFE α Bpa cross-linking in preinitiation complexes was tested using 1 nM site-specifically radiolabelled *gdh* promoter DNA, at 10 nM RNAP, 60 nM TBP, 120 nM TFB1, and 500 nM TFE α . Radiolabelled *gdh* promoter templates were made from biotinylated PCR products attached to paramagnetic streptavidin beads. The non-biotinylated strand was dissociated, and primers annealed and extended with radiolabelled dATP, washed, and fully extended with 500 μM unlabeled dNTPs. The fully extended products were released by restriction enzyme digestion, phenol extracted and ethanol precipitated, and then dissolved and quantified by scintillation counting. Reactions were assembled in 25 microliters overlaid with mineral oil, incubated for 30 min at 65 °C, and cross-links were formed by direct illumination with a handheld UV lamp (365 nm) at 2 cm distance for 60 min. DNA was digested with DNaseI and micrococcal nuclease (10 units of each), and samples were loaded on 4–20% gradient polyacrylamide gels⁵⁵. See Supplementary Fig. 8 for full scans of the gels.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The atomic structures of native and SeMet substituted *Pfu* TFE α Δ ZRD and their structure factors were deposited in the Protein Data Bank with the accession code 6PLN and 6XJF, respectively. The cryo-EM maps of *Tko* apo-RNAP, RNAP-TFE α complex, and RNAP-TFE α -DNA complex were deposited in the Electron Microscopy Data Bank with the accession codes EMD-9960, EMD-9961, and EMD-9962, respectively. Their corresponding models were deposited in the Protein Data Bank with the accession codes 6KF3, 6KF4, and 6KF9, respectively. Other data are available upon request. PDB files used for structure comparisons (Supplementary Figs. 5, 6, and 9: 5GPY, 3P8B, 5XON, 5IYA, 5FZ5, and 5FMF) are from the Protein Data Bank (<https://www.rcsb.org>).

Received: 27 April 2020; Accepted: 29 October 2020;

Published online: 30 November 2020

References

- Hirata, A., Klein, B. J. & Murakami, K. S. The X-ray crystal structure of RNA polymerase from Archaea. *Nature* **451**, 851–854 (2008).
- Jun, S. H. et al. The X-ray crystal structure of the euryarchaeal RNA polymerase in an open-clamp configuration. *Nat. Commun.* **5**, 5132 (2014).

3. Korkhin, Y. et al. Evolution of complex RNA polymerases: the complete archaeal RNA polymerase structure. *PLoS Biol.* **7**, e1000102 (2009).
4. Hahn, S. Structure and mechanism of the RNA polymerase II transcription machinery. *Nat. Struct. Mol. Biol.* **11**, 394–403 (2004).
5. Jun, S. H., Reichlen, M. J., Tajiri, M. & Murakami, K. S. Archaeal RNA polymerase and transcription regulation. *Crit. Rev. Biochem. Mol. Biol.* **46**, 27–40 (2011).
6. Grunberg, S. & Hahn, S. Structural insights into transcription initiation by RNA polymerase II. *Trends Biochem. Sci.* **38**, 603–611 (2013).
7. Plaschka, C. et al. Transcription initiation complex structures elucidate DNA opening. *Nature* **533**, 353–358 (2016).
8. He, Y. et al. Near-atomic resolution visualization of human transcription promoter opening. *Nature* **533**, 359–365 (2016).
9. Murakami, K. et al. Structure of an RNA polymerase II preinitiation complex. *Proc. Natl Acad. Sci. USA* **112**, 13543–13548 (2015).
10. Dienemann, C., Schwab, B., Schilbach, S. & Cramer, P. Promoter distortion and opening in the RNA polymerase II cleft. *Mol. Cell* **73**, 97–106.e4 (2019).
11. Schulz, S. et al. TFE and Spt4/5 open and close the RNA polymerase clamp during the transcription cycle. *Proc. Natl Acad. Sci. USA* **113**, E1816–E1825 (2016).
12. Werner, F. & Grohmann, D. Evolution of multisubunit RNA polymerases in the three domains of life. *Nat. Rev. Microbiol.* **9**, 85–98 (2011).
13. Vannini, A. & Cramer, P. Conservation between the RNA polymerase I, II, and III transcription initiation machineries. *Mol. Cell* **45**, 439–446 (2012).
14. Sarmiento, F., Mrazek, J. & Whitman, W. B. Genome-scale analysis of gene function in the hydrogenotrophic methanogenic archaeon *Methanococcus maripaludis*. *Proc. Natl Acad. Sci. USA* **110**, 4726–4731 (2013).
15. Meinhart, A., Blobel, J. & Cramer, P. An extended winged helix domain in general transcription factor E/II α . *J. Biol. Chem.* **278**, 48267–48274 (2003).
16. Grohmann, D. et al. The initiation factor TFE and the elongation factor Spt4/5 compete for the RNAP clamp during transcription initiation and elongation. *Mol. Cell* **43**, 263–274 (2011).
17. Najji, S., Grunberg, S. & Thomm, M. The RPB7 orthologue E' is required for transcriptional activity of a reconstituted archaeal core enzyme at low temperatures and stimulates open complex formation. *J. Biol. Chem.* **282**, 11047–11057 (2007).
18. Hanzelka, B. L., Darcy, T. J. & Reeve, J. N. TFE, an archaeal transcription factor in *Methanobacterium thermoautotrophicum* related to eucaryal transcription factor TFII α . *J. Bacteriol.* **183**, 1813–1818 (2001).
19. Grunberg, S., Bartlett, M. S., Najji, S. & Thomm, M. Transcription factor E is a part of transcription elongation complexes. *J. Biol. Chem.* **282**, 35482–35490 (2007).
20. Kuldell, N. H. & Buratowski, S. Genetic analysis of the large subunit of yeast transcription factor IIE reveals two regions with distinct functions. *Mol. Cell Biol.* **17**, 5288–5298 (1997).
21. Chen, H. T., Warfield, L. & Hahn, S. The positions of TFIIIF and TFIIIE in the RNA polymerase II transcription preinitiation complex. *Nat. Struct. Mol. Biol.* **14**, 696–703 (2007).
22. Grunberg, S., Warfield, L. & Hahn, S. Architecture of the RNA polymerase II preinitiation complex and mechanism of ATP-dependent promoter opening. *Nat. Struct. Mol. Biol.* **19**, 788–796 (2012).
23. Miwa, K. et al. Crystal structure of human general transcription factor TFIIIE at atomic resolution. *J. Mol. Biol.* **428**, 4258–4266 (2016).
24. Walker, J. E. & Santangelo, T. J. Analyses of in vivo interactions between transcription factors and the archaeal RNA polymerase. *Methods* **86**, 73–79 (2015).
25. Mekler, V., Minakhin, L. & Severinov, K. A critical role of downstream RNA polymerase-promoter interactions in the formation of initiation complex. *J. Biol. Chem.* **286**, 22600–22608 (2011).
26. Klein, B. J. et al. RNA polymerase and transcription elongation factor Spt4/5 complex structure. *Proc. Natl Acad. Sci. USA* **108**, 546–550 (2011).
27. Feklistov, A. & Darst, S. A. Structural basis for promoter-10 element recognition by the bacterial RNA polymerase sigma subunit. *Cell* **147**, 1257–1269 (2011).
28. Ryu, Y. & Schultz, P. G. Efficient incorporation of unnatural amino acids into proteins in *Escherichia coli*. *Nat. Methods* **3**, 263–265 (2006).
29. Dorman, G. & Prestwich, G. D. Benzophenone photophores in biochemistry. *Biochemistry* **33**, 5661–5673 (1994).
30. Micorescu, M. et al. Archaeal transcription: function of an alternative transcription factor B from *Pyrococcus furiosus*. *J. Bacteriol.* **190**, 157–167 (2008).
31. Bartlett, M. S., Thomm, M. & Geiduschek, E. P. Topography of the euryarchaeal transcription initiation complex. *J. Biol. Chem.* **279**, 5894–5903 (2004).
32. Hirata, A. et al. Archaeal RNA polymerase subunits E and F are not required for transcription in vitro, but a *Thermococcus kodakarensis* mutant lacking subunit F is temperature-sensitive. *Mol. Microbiol.* **70**, 623–633 (2008).
33. He, Y., Fang, J., Taatjes, D. J. & Nogales, E. Structural visualization of key steps in human transcription initiation. *Nature* **495**, 481–486 (2013).
34. Hoffmann, N. A. et al. Molecular structures of unbound and transcribing RNA polymerase III. *Nature* **528**, 231–236 (2015).
35. Kostrewa, D. et al. RNA polymerase II-TFIIB structure and mechanism of transcription initiation. *Nature* **462**, 323–330 (2009).
36. Kettenberger, H., Armache, K. J. & Cramer, P. Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Mol. Cell* **16**, 955–965 (2004).
37. Feklistov, A. et al. RNA polymerase motions during promoter melting. *Science* **356**, 863–866 (2017).
38. Boyaci, H., Chen, J., Jansen, R., Darst, S. A. & Campbell, E. A. Structures of an RNA polymerase promoter melting intermediate elucidate DNA unwinding. *Nature* **565**, 382–385 (2019).
39. Chakraborty, A. et al. Opening and closing of the bacterial RNA polymerase clamp. *Science* **337**, 591–595 (2012).
40. Bae, B. et al. CarD uses a minor groove wedge mechanism to stabilize the RNA polymerase open promoter complex. *Elife* **4**, e08505 (2015).
41. Davis, E., Chen, J., Leon, K., Darst, S. A. & Campbell, E. A. Mycobacterial RNA polymerase forms unstable open promoter complexes that are stabilized by CarD. *Nucleic Acids Res* **43**, 433–445 (2015).
42. Doublet, S. Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol.* **276**, 523–530 (1997).
43. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
44. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 213–221 (2010).
45. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D. Biol. Crystallogr.* **60**, 2126–2132 (2004).
46. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
47. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
48. Kimanius, D., Forsberg, B. O., Scheres, S. H. & Lindahl, E. Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. *Elife* **5**, e18722 (2016).
49. Tang, G. et al. EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).
50. Scheres, S. H. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nat. Methods* **9**, 853–854 (2012).
51. Vilas, J. L. et al. MonoRes: automatic and accurate estimation of local resolution for electron microscopy maps. *Structure* **26**, 337–344.e4 (2018).
52. Afonine, P. V. et al. New tools for the analysis and validation of cryo-EM maps and atomic models. *Acta Crystallogr. D. Struct. Biol.* **74**, 814–840 (2018).
53. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
54. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 12–21 (2010).
55. Dext, S. et al. Displacement of the transcription factor B reader domain during transcription initiation. *Nucleic Acids Res.* **46**, 10066–10081 (2018).
56. Waegel, I., Schmid, G., Thumann, S., Thomm, M. & Hausner, W. Shuttle vector-based transformation system for *Pyrococcus furiosus*. *Appl. Environ. Microbiol.* **76**, 3308–3313 (2010).

Acknowledgements

We thank the staff at the Jinju Bio-industry Foundation for supporting *Tko* cell fermentation and Macromolecular X-ray science at the Cornell High Energy Synchrotron Source (MacCHESS) for supporting the crystallographic data collection. The cryo-EM experiment was performed at the Cryo-EM facility in Electron Microscopy Research Center, Korea Basic Science Institute (KBSI) and supported by KBSI grant C030221 and C021820. We thank Carmen V. Amoah-Kusi for assistance with cross-linking and transcription assays. This work was supported by the Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Education (NRF-2015R1D1A1A01059097 to S.-H.J.) and by the Ministry of Science, ICT & Future Planning (NRF-2016R1A5A1010764 and NRF-2017M3A9F6029755 to H.-S.C. and NRF-2020R1F1A1072050 to S.-H.J.) and NIH grants (R01 GM087350 and R35 GM131860 to K.S.M. and R15 GM083306 to M.S.B.).

Author contributions

K.S.M., S.-H.J., M.S.B., and H.-S.C. conceived and designed the experiments. S.-H.J. and K.S.M. crystallized and solved TFE α crystal structure. S.-H.J., J.S.C., and H.K. prepared *Tko* RNAP and its complexes for cryo-EM. J.H. conducted cryo-EM sample preparation, data collection, and image processing. S.-H.J. and K.S.M. carried out cryo-EM model building. M.S.B. performed cross-linking and transcription assays. K.S.M.,

S.-H.J., H.-S.C., J.H., and M.S.B. wrote the paper. All authors discussed the results and commented on the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41467-020-19998-x>.

Correspondence and requests for materials should be addressed to S.-H.J., H.-S.C. or K.S.M.

Peer review information *Nature Communications* thanks Finn Werner and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020