Transcription activation of bacteriophage T4 late genes is accomplished by a transcription activation complex containing RNA polymerase (RNAP), the promoter specificity factor gp55, the coactivator gp33, and a universal component of cellular DNA replication, the sliding clamp gp45. Although genetic and biochemical studies have elucidated many aspects of T4 late gene transcription, no precise structure of the transcription machinery in the process is available. Here, we report the cryo-EM structures of a gp55-dependent RNAP-promoter open complex and an intact gp45-dependent transcription activation complex. The structures reveal the interactions between gp55 and the promoter DNA that mediate the recognition of T4 late promoters. In addition to the $\sigma$R2 homology domain, gp55 has a helix-loop-helix motif that chaperones the template-strand single-stranded DNA of the transcription bubble. Gp33 contacts both RNAP and the upstream double-stranded DNA. Gp45 encircles the DNA and tethers RNAP to it, supporting the idea that gp45 switches the promoter search from three-dimensional diffusion mode to one-dimensional scanning mode.
To initiate the transcription of most genes, E. coli RNA polymerase (RNAP, subunit composition $\alpha_2\beta\beta^{\prime}\omega$) forms holoenzyme with $\sigma^{70}$, which contacts RNAP extensively and recognizes the promoter DNA\textsuperscript{1-7}. Specifically, $\sigma$ conserved region $\sigma R2$ contacts the clamp helices of the RNAP $\beta'$ subunit and recognizes the promoter -10 element, while $\sigma$ conserved region $\sigma R4$ contacts the flap tip helix (FTH) of the RNAP $\beta$ subunit and recognizes the promoter -35 element.

Bacteriophages use bacterial RNAP to transcribe their own genes. For decades, transcription of bacteriophage T4 late genes has served as a model system to investigate mechanisms of transcription regulation. The T4 late promoters consist of a -10-like element placed at the position corresponding to the bacterial promoter -10 element\textsuperscript{8}. Nevertheless, there is no sequence conservation at the position corresponding to the bacterial promoter -35 element. Herendeen et al. developed an in vitro system that leads to the current understanding of T4 transcription, bacterial RNAP needs to form holoenzyme with gp55, gp33, gp45, and the nucleic acid scaffold, freeze the sample, and collect data on Titan Krios. Steps of 3D classification reveal the structures of RPo-gp55 and TAC-gp45 (Fig. 1, Supplementary Figs. 2–5, and Supplementary Table 1). The structure of RPo-gp55 is determined at a nominal resolution of 3.6 Å (Supplementary Fig. 3a). Local resolution calculation indicates that the central core of the structure is determined to 3.0–4.0 Å resolution (Supplementary Fig. 3c). The structure of TAC-gp45 is determined at a nominal resolution of 4.5 Å (Supplementary Fig. 4a). Local resolution calculation indicates that the central core of the structure is determined to be 4.0–6.0 Å (Supplementary Fig. 4c).

Experimental density maps show clear densities for RNAP, gp55, gp33, gp45, and the nucleic acid scaffold (Fig. 1b, c and Supplementary Fig. 5). The RNAPs of the structures are very similar to the previously reported $\sigma^{70}$-dependent RNAP-promoter open complex (RPo-g$\sigma^{70}$) structure\textsuperscript{6}. Consistent with the previous KMMo\textsubscript{4} footprinting experiments\textsuperscript{13}, nucleic acid scaffold from -12 to +2 is unwound in RPo-gp55 and TAC-gp45, while the upstream dsDNA and the downstream dsDNA adopt similar orientations as in RPo-g$\sigma^{70}$. gp55 adopts the same conformation and makes essentially identical interactions with RNAP and the promoter DNA in RPo-gp55 and TAC-gp45. The interactions in RPo-gp55 will be discussed in the following sections due to its superior resolution.

**Results**

Cryo-EM structures of RPo-gp55 and TAC-gp45. E. coli RNAP, gp55, gp33, gp45, clamp loader gp44/gp62, and single-stranded DNA (ssDNA)-binding protein gp32 are purified to homogeneity (Supplementary Fig. 1a). To ascertain the activity of the purified proteins, we develop an in vitro transcription assay by taking advantage of a fluorogenic RNA aptamer, Mango III (Supplementary Fig. 1b)\textsuperscript{21}. The nucleic acid scaffold encompasses the essential features of a T4 late transcription unit that can be activated by gp45 (Fig. 1a). It includes the T4 late promoter P\textsubscript{23} followed by Mango III encoding sequence and rnb terminators (T1 and T2). Loading of gp45 onto this DNA with the appropriate polarity for transcription activation is assured by ~150 nt of 5’ overhanging ssDNA, generated by exonuclease III, at the downstream DNA end only. The transcription activities of RNAP in the presence of gp55, gp33, and gp45 are evaluated by measuring the fluorescence intensities of Mango III (Supplementary Fig. 1c).

To obtain a structure of the T4 late gene transcription machinery, we incubate E. coli RNAP, gp55, gp33, gp44, gp45/gp62, gp32, and the nucleic acid scaffold, freeze the sample, and collect data on Titan Krios. Steps of 3D classification reveal the structures of RPo-gp55 and TAC-gp45 (Fig. 1, Supplementary Figs. 2–5, and Supplementary Table 1). The structure of RPo-gp55 is determined at a nominal resolution of 3.6 Å (Supplementary Fig. 3a). Local resolution calculation indicates that the central core of the structure is determined to 3.0–4.0 Å resolution (Supplementary Fig. 3c). The structure of TAC-gp45 is determined at a nominal resolution of 4.5 Å (Supplementary Fig. 4a). Local resolution calculation indicates that the central core of the structure is determined to be 4.0–6.0 Å (Supplementary Fig. 4c).

Experimental density maps show clear densities for RNAP, gp55, gp33, gp45, and the nucleic acid scaffold (Fig. 1b, c and Supplementary Fig. 5). The RNAPs of the structures are very similar to the previously reported $\sigma^{70}$-dependent RNAP-promoter open complex (RPo-g$\sigma^{70}$) structure\textsuperscript{6}. Consistent with the previous KMMo\textsubscript{4} footprinting experiments\textsuperscript{13}, nucleic acid scaffold from -12 to +2 is unwound in RPo-gp55 and TAC-gp45, while the upstream dsDNA and the downstream dsDNA adopt similar orientations as in RPo-g$\sigma^{70}$. gp55 adopts the same conformation and makes essentially identical interactions with RNAP and the promoter DNA in RPo-gp55 and TAC-gp45. The interactions in RPo-gp55 will be discussed in the following sections due to its superior resolution.

$\sigma R2$ homology domain of gp55 contacts the clamp helices. Gp55 is composed of five helices (H1–H5) connected by short linkers (Fig. 2a). As predicted by the sequence analysis\textsuperscript{10}, the folding of H1–H4 is reminiscent of $\sigma R2$ (Fig. 2b, c and Supplementary Fig. 5a). However, H4 is much longer than the equivalent helix in $\sigma R2$ and forms a helix-loop-helix (HLH) motif, together with H5 and their linker.

There is no clear density for two terminal segments of gp55. In an alignment of gp55 homologs, the N-terminal region is poorly conserved, and many homologs lack this region entirely\textsuperscript{23}. The C-terminal region of gp55 contains the SCBM, which would not be expected to interact with RNAP or DNA, but instead to be available for interaction with gp45. Accordingly, the C-terminal region truncated gp55 binds RNAP, forms RPo-gp55, and
Fig. 1 Cryo-EM structures of RPo-gp55 and TAC-gp45. **a** Nucleic acid scaffold used for cryo-EM. The nucleic acid scaffold encompasses the T4 late promoter $P_{23}$ followed by Mango III encoding sequence and $rrnB$ terminators ($T_1$ and $T_2$). Loading of gp45 onto this DNA with the appropriate polarity for transcription activation is assured by ~150 nt of 5' overhanging ssDNA at the downstream DNA end only. Salmon, nontemplate-strand DNA; red, template-strand DNA; blue, $P_{23}$; green, Mango; cyan, $T_1$ and $T_2$. Positions are numbered relative to the transcription start site. **b** The cryo-EM density map and the superimposed model of RPo-gp55. Light green, RNAP $\alpha$ subunits; wheat, RNAP $\beta$ subunit; light pink, RNAP $\beta'$ subunit; yellow, gp55; salmon, nontemplate-strand DNA; red, template-strand DNA; blue, -10-like element; dsDNA, double-stranded DNA. **c** The cryo-EM density map and the superimposed model of TAC-gp45. Dark green, gp33; cyan, gp45I; turquoise, gp45II; blue, gp45III. View orientation and other colors as in b.
initiates basal transcription as effectively as full-length gp55 (Supplementary Fig. 6).

The σR2 homology domain of gp55 binds to the clamp helices of RNAP. In particular, three negatively charged residues (E70, D74, E77) of H3 make electrostatic interactions with a positively charged groove of the clamp helices (Fig. 2d and Supplementary Fig. 5b), consistent with the previous report that amino acid substitutions of these residues strongly disrupt the gp55–RNAP interaction in competitive binding experiments.\(^{12}\)

σR2 homology domain of gp55 mediates the recognition of the promoter -10-like element. The σR2 homology domain of gp55 mediates the recognition of the promoter -10-like element through interactions with nontemplate-strand ssDNA in the unwound transcription bubble (Fig. 3a and Supplementary Fig. 5c). The σR2 homology domain of gp55 unstacks nucleotides, flips nucleotides, and inserts nucleotides into protein pockets at positions -12, -9, and -8. In particular, the adenine at position -12 is inserted into a deep protein pocket, making a stacking

Fig. 2 The σR2 homology domain of gp55 contacts the clamp helices. a Structural organization of gp55. Two terminal segments are disordered and colored gray. Observed secondary structure features (H1–H5) are shown schematically and labeled. Yellow, σR2 homology domain; orange, helix-loop-helix (HLH) motif; SCBM, sliding clamp-binding motif. b The σR2 homology domain contacts the clamp helices. Yellow, σR2 homology domain; orange, helix-loop-helix (HLH) motif; light pink, clamp helices, zipper, and lid. c σR2 in RPo-σ\(^{70}\) (PDB 6CA0 [https://doi.org/10.2210/pdb6CA0/pdb]). View orientation as in b. d Three negatively charged residues of H3 make electrostatic interactions with a positively charged groove of the clamp helices. Blue represents positively charged surfaces (\(+5\) kT) and red represents negatively charged surfaces (\(-5\) kT).

Fig. 3 The σR2 homology domain of gp55 mediates the recognition of the promoter -10-like element. a Interactions of gp55 with nontemplate-strand ssDNA. Gp55 is shown as surfaces and DNA is shown as sticks. b For promoter position -12, the base is unstacked and inserted into a protein pocket. c For promoter positions -9 and -8, the bases are unstacked and inserted into protein pockets. d Electrophoretic mobility shift assays show that substitutions of the pocket residues impair RPo-gp55 formation. Experiments were repeated independently three times with similar results. e Mango III transcription assays show that substitutions of the pocket residues jeopardize gp55-dependent transcription. Error bars represent mean ± SEM out of \(n = 3\) experiments.
interaction with residue Y97 (Fig. 3b). The adenine at position -9 is inserted into a shallow protein pocket, making van der Waals interactions with residues N13 and H95, while the adenine at position -8 is inserted into a deep protein pocket, making a H-bonded interaction with residue T99 (Fig. 3c). Alanine substitutions of residues (N13, L54, R57, F86, H95, Y97, and T99) in these protein pockets impair RPo-gp55 formation and gp55-dependent transcription (Fig. 3d, e), verifying that the cryo-EM structures are biologically relevant.

HLH motif of gp55 chaperons the template-strand ssDNA of the transcription bubble. The HLH motif of gp55 contacts the zipper, displaces the lid of RNAP, and chaperons the template-strand ssDNA of the transcription bubble. Gray, lid in RPo-σ70 (PDB 6CA0 [https://doi.org/10.2210/pdb6CA0/pdb]); red, template-strand ssDNA. View orientation and other colors as in Fig. 2b. a The HLH motif of gp55 contacts the zipper, displaces the lid of RNAP, and chaperons the template-strand ssDNA of the transcription bubble. Gray, lid in RPo-σ70 (PDB 6CA0 [https://doi.org/10.2210/pdb6CA0/pdb]); red, template-strand ssDNA. View orientation and other colors as in Fig. 2b. b Gp55 residue R114 makes an electrostatic interaction with the DNA backbone phosphate at position -11. Colors as in a. c Electrophoretic mobility shift assays show that both truncation of the HLH motif and substitution of residue R114 impair RPo-gp55 formation. Experiments were repeated independently three times with similar results. d Mango III transcription assays show that both truncation of the HLH motif and substitution of residue R114 jeopardize gp55-dependent transcription. Error bars represent mean ± SEM out of n = 3 experiments.
substitution of E64 and E65 in combination decreases gp33–RNAP interaction in competitive binding experiments15.

Interactions between gp33 and the upstream dsDNA. In our structure, the H3 of gp33 contacts the upstream dsDNA. In particular, residue K75 is positioned to form a salt bridge with the DNA backbone phosphate (Fig. 5c and Supplementary Fig. 5f). Gp33 mutant protein (K75A) represses basal transcription as efficiently as wild-type protein (Fig. 5d), indicating that the mutant protein is well folded and the gp33–DNA interaction is not required for transcription repression. However, the mutant protein activates transcription to a less extent than wild-type protein, indicating that the gp33–DNA interaction is important for transcription activation. In addition, gp33 contacts the DNA backbone only, consistent with the observation that the sequences of the upstream dsDNA are not conserved in T4 late promoters23.

Interactions between gp45 and the SCBMs. A triangular density feature adjacent to gp33 can be attributed to gp45 homotrimer (Supplementary Fig. 5g). For convenience, the three protomers (gp45I, gp45II, and gp45III) are numbered clockwise, looking from downstream to upstream (Fig. 6a). The upstream dsDNA goes through the center of the closed ring of gp45. The lateral face with the protruding C-terminus presents a hydrophobic patch on each protomer that serves as a binding site for the SCBMs of DNA polymerase, DNA ligase, gp33, and gp5518,24. Unfortunately, the local resolution of gp45 is too low to build the model for the SCBMs of gp33 and gp55 (Supplementary Fig. 4c). To determine which protomers gp33 and gp55 most likely bind to, we superimpose the crystal structure of gp45 in complex with the SCBM of T4 DNA ligase24 on the cryo-EM structure of TAC-gp45. The distance between the C-terminus of the gp33 structure (residue 102) and the N-terminus of the SCBM (corresponding to residue 106 of gp33 according to Supplementary Table 2) bound to gp45I is 13 Å, which is within the range of distances that could be spanned by a three-residue linker. Therefore, the SCBM of gp33 probably binds to gp45I. The distances between the C-terminus of the gp55 structure (residue 153) and the N-termini of the SCBMs (corresponding to residue 175 of gp55 according to Supplementary Table 2) bound to gp45II and gp45III are 45 and 71 Å, respectively, which are within the range of distances that could be spanned by a 21-residue linker, but the linker needs to bypass the upstream dsDNA in the case of gp45III (Fig. 6b). Therefore, the SCBM of gp55 probably binds to gp45II.

Discussion
In this work, we determine the precise structures of T4 late gene transcription machinery. The structures unequivocally show that...
a sliding clamp can turn on transcription by physically tethering RNAP to the DNA. Our structures confirm that gp55 is a diverged σ70 family protein, contacting the clamp helices and mediating the sequence-specific interactions with the promoter -10-like element. In addition to the σ2 homology domain, we find that gp55 has a HLH motif, which is critical for T4 late transcription. The structure of TAC-gp45 reveals that gp33 contacts the upstream dsDNA, which is important for gp45-dependent transcription activation. The structure verifies that gp45 is located at the upstream end of TAC-gp45, in the vicinity of gp33. The gp45 protomers, to which gp33 and gp55 bind, are assigned based on the restraints of their linkers.

A model of gp45-dependent transcription activation has been proposed based on the elegant biochemical studies by Geiduschek and Kassavetis (Fig. 7). To initiate transcription, RNAP must locate promoter sequences, which compose a very small amount of the genome. Single-molecule experiments have demonstrated that the promoter search is dominated by three-dimensional diffusion. As for bacteriophage T4 late gene transcription, ATP hydrolysis by gp44/gp62 complex triggers the detachment of gp45 from the gp44/gp62 complex, freeing gp45 to track along DNA. RNAP–gp55–gp33 grabs gp45 through the SCBMs of gp33 and gp55, and travels along with gp45. This kind of one-dimensional search is intuitively more efficient than three-dimensional diffusion.

The cryo-EM structure of TAC-gp45 shows that the HLH motif of gp55 blocks the exit path for the nascent RNA (Supplementary Fig. 7) and must be displaced to allow further extension of the RNA and transition into transcription elongation complex. After gp55 is displaced, the general transcription factor NusG could bind to the clamp helices and enhance elongation by suppressing RNAP backtracking. The cryo-EM structure further shows that gp33 and gp45 would not occlude the RNA exit channel and could remain bound to RNAP during elongation. Even though gp55 may be displaced from RNAP, it could remain tethered to gp45 through the SCBM. Therefore, the quaternary complex resumes promoter scanning immediately after transcription termination, which may further boost T4 late gene transcription. The binding of gp33 during elongation would decrease the accessibility of the FTH to the general transcription factor NusA, which plays key roles in σ-dependent transcription termination and transcription-translation coupling. Therefore, both processes would be compromised during T4 late transcription. Although our structures are consistent with the model that the quaternary complex stays together during elongation, we cannot rule out the possibility that gp33, gp55, and gp45 might fall off RNAP during elongation.

Transcription activation is a ubiquitous strategy of transcription regulation and it generally involves activators. Most activators are DNA-binding proteins that bind to enhancers or promoter-proximal elements. Instead of using DNA-binding activators, T4 bacteriophage activates the transcription of late genes by using a DNA-tracking protein, the sliding clamp, as the activator. As a proof of concept, fusion proteins linking the transcription activation domain of herpes simplex virus VP16 protein to a sliding clamp have been shown to activate transcription by yeast and Drosophila RNAII in vitro. It is likely that other instances of the use of this simple mechanism for transcription activation would be found in nature.

**Methods**

**Gp55.** Gp55 was prepared as described. Briefly, E. coli strain BL21(DE3) (Invitrogen, Inc.) was transformed with plasmid pET21a-gp55 (GENEWIZ, Inc.) encoding gp55 under the control of the bacteriophage T7 gene 10 promoter. Single colonies of the resulting transformants were used to inoculate 11 LB broth containing 100 μg/ml ampicillin, cultures were incubated at 37 °C with shaking until OD600 = 0.9, cultures were induced by addition of IPTG to 1 mM, and cultures were incubated 17 h at 16 °C. Then, cells were harvested by centrifugation (5,422 × g; 15 min at 4 °C), resuspended in 20 ml buffer A (50 mM Tris-Cl, pH 7.9, 0.1 M NaCl, and 5% glycerol), and lysed using a JN-02C cell disrupter (NJBIO, Inc.). After centrifugation (24,792 × g; 40 min at 4 °C), the pellet was washed with 25 ml.
buffer A supplemented with 0.25% Triton X-100 twice, resuspended in buffer B (50 mM Tris-HCl, pH 7.9, 6 mM GtHCl, 1 mM EDTA, and 10% glycerol), and dialyzed against 1 ml buffer C (20 mM Tris-HCl, pH 7.4, 1 mM EDTA, and 5 mM β-mercaptoethanol) for three times. After centrifugation (24,792 × g; 30 min at 4 °C), the supernatant was dialyzed against buffer D (20 mM Tris-HCl, pH 7.9, 1 mM EDTA, 1 mM DTT, and 5% glycerol); loaded onto a Mono Q 10/100 GL column (GE Healthcare, Inc.) equilibrated in buffer D, and eluted with a 160 ml linear gradient of 0–1 M NaCl in buffer D. The column was eluted to 5 ml using an Amicon Ultra-15 Centrifugal Filter (10 kDa MWCO; Merck Millipore, Inc.) and applied to a HiLoad 16/600 Superdex 200 column (GE Healthcare, Inc.) equilibrated in buffer E (10 mM HEPES, pH 7.5, 0.05 M KCl), and the column was eluted with 120 ml of the same buffer.

**E. coli RNAP.** E. coli RNAP was prepared from E. coli strain BL21(DE3) (Invitrogen, Inc.) transformed with plasmid pLH240-35. Single colonies of the resulting transformants were used to inoculate 50 ml LB broth containing 100 μg/ml ampicillin, and cultures were incubated 16 h at 37 °C with shaking. Aliquots (10 ml) were used to inoculate 11 LB broth containing 100 μg/ml ampicillin, cultures were incubated at 37 °C with shaking until OD600 = 0.6, cultures were induced by addition of IPTG to 1 mM, and cultures were incubated 3 h at 37 °C. Then, cells were harvested by centrifugation (5422 × g; 15 min at 4 °C), resuspended in 20 ml lysis buffer (20 mM Tris-HCl, pH 8.0, 0.5 M NaCl, 5% glycerol), and lysed using a JN-02C cell disruptor (INBio, Inc.). The lysate was centrifuged (24,792 × g; 40 min at 4 °C), and the supernatant was dialyzed against 1 ml buffer containing 0.4 M imidazole and eluted with 10 ml lysis buffer containing 0.16 M imidazole. The sample was further purified by anion-exchange chromatography on a Mono Q 10/100 GL column (GE Healthcare, Inc.; 160 ml linear gradient of 0.1 M NaCl in buffer D). Fractions containing gp33 were concentrated to 5 ml using an Amicon Ultra-15 centrifugal filter (10 kDa MWCO; Merck Millipore, Inc.) and applied to a HiLoad 16/600 Superdex 200 column (GE Healthcare, Inc.) equilibrated in buffer E, and the column was eluted with 120 ml of the same buffer.

**Nuclease scaffold.** Double-stranded, 3′-end recessed DNA scaffolds were prepared by Exo III treatment as described13. Briefly, an ~600-bp PCR fragment containing the B2 promoter, the mango III coding sequence, and rnr8 terminators amplified from plasmid pUC57-P2-gp-EX (MangoGeneWiz, Inc.) using primers in Supplementary Table 3 was cleaved with Hind III and Xho I to generate Exo III-susceptible downstream and Exo III-resistant upstream ends and reacted with Exo III to generate ~150 nt of ssDNA.

**Mango III transcription assay.** Mango III transcription assay was performed in a 384-well microplate format. Reaction mixtures contained (50 μl): 0.2 μM E. coli RNAP, 0.02 μM nucleic acid scaffold, 0 or 0.4 μg gp55, 0 or 0.4 μg gp44 homotrimer, 0.35 μg gp32, 2 mM dATP, 0.1 mg/ml heparin, 1 mM ATP, 1 mM GTP, 1 mM UTP, 1 mM CTP, and 1 mM MgCl2 in transcription buffer, then plunge-frozen in liquid ethane using a Vitrobot (FEI, Inc.) with 95% chamber humidity at 10 °C.

**Cryo-EM data acquisition and processing.** Cryo-EM grid preparation. Quantifoil grids (R 1.2/1.3 Cu 300 mesh; Quantifoil, Inc.) were glow-discharged for 120 s at 25 mA prior to the application of 3 μl of the complex in transcription buffer, then plunged-frozen in liquid ethane using a Vitrobot (FEI, Inc.) with 95% chamber humidity at 10 °C.

**Cryo-EM data acquisition and processing.** The grids were imaged using a 300 kV Titan Krios (FEI, Inc.) equipped with a K2 Summit direct electron detector (Gatan, Inc.). Images were recorded with Serial EM36 in counting mode with a physical pixel size of 1.307 Å and a defocus range of 1.5-2.5 μm. Data were collected with a dose of 10 e⁻/Å²/s. Images were recorded with a 10 e⁻ exposure and 0.25 s sub-frames to give a total dose of 59 e⁻/Å². Subframes were aligned and summed using MotionTrack. The contrast transfer function was estimated using cifCTFIND37. From the summed images, ~10,000 particles were manually picked and subjected to 2D classification in RELION5. 2D averages of the best classes were used as templates for auto-picking in RELION. Auto-picked particles were manually inspected, then subjected to 2D classification in RELION. The resulting 3D classes were selected by averaging the stacks and selecting the 3D class with the highest sum of EM images using a map of E. coli TEC (EMD-8855)38 in RELION using a map of E. coli TEC (EMD-8855) [https://www.emdatresource.org/EMD-8855] low-pass filtered to 40 Å resolution as a reference. Initial 3D
classification resulted in a reconstruction with densities for gp45 (TAC-gp55) and a reconstruction without densities for gp45 (RPo-gp55). For RPo-gp55, further 3D classification with four classes was performed. The density map of the upstream dsDNA of class 1 is weak compared with the other three classes. Therefore, only reconstruction without densities for gp45 (RPo-gp55). For RPo-gp55, further 3D classification with four classes was performed. The density map of the upstream dsDNA of class 1 is weak compared with the other three classes. Therefore, only

Cryo-EM model building and refinement. The homology model of gp55 was generated on Phyre2 server41. The model of RNP from the structure of RPo-gp55 (PDB 6CA0 [https://doi.org/10.2210/pdb6CA0/pdb]), the homology model of gp55, and the crystal structures of gp33 and gp45 were fitted into the cryo-EM density map using Chimera42. The coordinates of nucleic acids were built manually in Coot43. The coordinates were real-space refinement and postprocessing.

Pull-down assay. Pull-down assays were performed, taking advantage of the fact that the σp subunit of E. coli RNAp is C-terminally His-tagged. Overall, 250 μg E. coli RNAp (1 μM) in buffer F was incubated with 50 μl Ni-NTA agarose (Qiagen, Inc.) for 5 min at 22 °C. After the resin was centrifuged (9391 g, 2 min at 22 °C) and washed with 0.5 ml buffer F for three times, bound proteins were analyzed by SDS-PAGE and quantified by ImageJ (https://imagej.nih.gov/ij/).

Electrophoretic mobility shift assay. Template-strand DNA (5'-AGTCCGGGCT GCTTCAGATCGAGATTTATCTCCGATATGTCGCTCTTT ATTAG-3', Sangon Biotech, Inc.) and non-template-strand DNA (5'-CTAAATAAAGGCCTCCAGATCTGGAATAATATCCGCTGATACG AGCCGGACT-3', Sangon Biotech, Inc.) were annealed at a 1:1 ratio in 10 mM Tris-HCl, pH 7.9, 0.2 M NaCl, and stored at ~80 °C. Electrophoretic mobility shift assays were performed in reaction mixtures containing (10 μl): 0.4 μM gp55 or gp55 derivatives, 0.2 μM E. coli RNAp, 0.2 μM DNA scaffold, 0.1 mg/ml heparin, and 5% glycerol in buffer E. E. coli RNAp was incubated with gp55 or gp55 derivatives for 10 min on ice, incubated with DNA scaffold for 20 min at 37 °C, and incubated with 0.1 mg/ml heparin for 1 min at 22 °C. The reaction mixtures were applied to 5% polyacrylamide slab gels (29:1 acrylamide/bisacrylamide), electrophoresed in 90 mM Tris-borate, pH 8.0, and 0.2 mM EDTA, stained with 4S Red Plus Nucleic Acid Stain (Sangon Biotech, Inc.) according to the procedure of the manufacturer.

References

1. Bae, B., Feklistov, A., Lasz-Napiorkowska, A., Landick, R. & Darst, S. A. Structure of a bacterial RNA polymerase holoenzyme open promoter complex. Elife 6, e28504 (2017).

2. Feklistov, A. & Darst, S. A. Structural basis for promoter–element recognition by the bacterial RNA polymerase sigma subunit. Cell 147, 1257–1269 (2011).

3. Feklistov, A., Sharon, B. D., Darst, S. A. & Gross, C. A. bacterial sigma factors: a historical, structural, and genomic perspective. Annu. Rev. Microbiol. 68, 375–403 (2014).

4. Feng, Y., Zhang, Y. & Eibright, R. H. Structural basis of transcription activation. Science 352, 1330–1333 (2016).

5. Hubin, E. A. et al. Structure and function of the mycobacterial transcription initiation complex with the essential regulator RbpA. Elife 6, e22520 (2017).

6. Narayan, A. et al. Cryo-EM structure of Escherichia coli sigma(70) RNA polymerase and promoter DNA complex revealed a role of sigma non-conserved region during the open complex formation. J. Biol. Chem. 293, 7367–7375 (2018).

7. Zhang, Y. et al. Structural basis of transcription initiation. Science 338, 1076–1080 (2012).

8. Christensen, A. C. & Young, E. T. T4 late transcripts are initiated near a conserved DNA sequence. Nature 299, 369–371 (1982).

9. Herendeen, D. R., Williams, K. P., Kassavetis, G. A. & Geiduschek, E. P. An RNA polymerase-binding protein that is required for communication between an enhancer and a promoter. Science 248, 573–578 (1990).

10. Grishkov, M. & Burgess, R. R. Sigma factors from E. coli, B. subtilis, phage SP01, and phage T4 are homologous proteins. Nucleic Acids Res. 14, 6745–6763 (1986).

11. Kassavetis, G. A., Elliott, T., Rabussay, D. P. & Geiduschek, E. P. Initiation of transcription at phase T4 late promoters with purified RNA polymerase. Cell 93, 887–897 (1993).

12. Nechaev, S., Kamali-Moghaddam, M., Andre, E., Leonetti, J. P. & Geiduschek, E. P. The bacteriophage T4 late-transcription coactivator gp33 binds the flap domain of Escherichia coli RNA polymerase. Proc. Natl Acad. Sci. USA 101, 17365–17370 (2004).

13. Kolesky, S. E., Ouhammouch, M. & Geiduschek, E. P. The mechanism of transcriptional activation by the topologically DNA-linked sliding clamp of bacteriophage T4. J. Mol. Biol. 321, 767–784 (2002).

14. Tinker, R. L., Sanders, G. M., Severinov, K., Kassavetis, G. A. & Geiduschek, E. P. The COOH-terminal domain of the RNA polymerase alpha subunit in transcriptional enhancement and deactivation at the bacteriophage T4 late promoter. J. Biol. Chem. 270, 15899–15907 (1995).

15. Shamoo, Y. & Steitz, T. A. Building a reposome from interacting pieces: sliding clamp complexed to a peptide from DNA polymerase and a polymerase editing complex. Cell 99, 155–166 (1999).

16. Sanders, G. M., Kassavetis, G. A. & Geiduschek, E. P. Dual targets of a transcriptional activator that tracks on DNA. EMBO J. 16, 3124–3132 (1997).

17. Wong, K. & Geiduschek, E. P. Activator-sigma interaction: a hydrophobic segment mediates the interaction of a sigma family promoter recognition protein with a sliding clamp transcription activator. J. Mol. Biol. 284, 195–203 (1998).

18. Shi, J. et al. Structural basis of Q-dependent transcription antitermination. Nat. Commun. 10, 2925 (2019).

19. Nechaev, S. & Geiduschek, E. P. Dissection of the bacteriophage T4 late promoter complex. J. Mol. Biol. 379, 402–413 (2008).

20. Geiduschek, E. P. & Kassavetis, G. A. Transcription of the T4 late genes. Virol. J. 7, 288 (2010).

21. Shi, K. et al. T4 DNA ligase structure reveals a prototypical ATP-dependent ligase with a unique mode of sliding clamp interaction. Nucleic Acids Res. 46, 10474–10488 (2018).

22. Friedman, L. J., Mumm, J. P. & Gelles, J. RNA polymerase approaches its promoter without long-range sliding along DNA. Proc. Natl Acad. Sci. USA 110, 9740–9745 (2013).

23. Wang, F. et al. The promoter-search mechanism of Escherichia coli RNA polymerase is dominated by three-dimensional diffusion. Nat. Struct. Mol. Biol. 20, 174–181 (2013).

24. Kang, J. Y. et al. Structural basis for transcript elongation control by nusG family universal regulators. Cell 173, 1650–1662 (2018).

25. Said, N. et al. Steps toward transcriptional-independent RNA polymerase inactivation by terminator ATPase rho. Science https://doi.org/10.1126/science.abd1673 (2021).

26. Wang, C. et al. Structural basis of transcription-translation coupling. Science 369, 1359–1365 (2020).

27. O’Reilly, F. J. et al. In-cell architecture of an actively transcribing-translating ribosome. Elife 12, e64677 (2013).

28. Brown, D. F. & Busby, S. J. Local and global regulation of transcription initiation in bacteria. Nat. Rev. Microbiol. 14, 708–720 (2016).

29. Osman, S. & Cramer, P. Structural biology of RNA polymerase II transcription: 20 years on. Annu. Rev. Cell Dev. Biol. 36, 1–34 (2020).
Acknowledgements

The authors would like to thank Shenghai Chang at the Center of Cryo Electron Microscopy in Zhejiang University School of Medicine for help with cryo-EM data collection. The authors would like to thank the Core Facilities, Zhejiang University School of Medicine, for the technical support. This work was funded by the National Natural Science Foundation of China (31970040 to Y.F. and 32000025 to J.S.).

Author contributions

J.S., A.W., S.I., B.G., and Y.H. performed the experiments. Y.F. supervised the experiments and wrote the manuscript with contributions from the other authors. All authors contributed to the analysis of the data and the interpretation of the results.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-021-21392-0.

Correspondence and requests for materials should be addressed to Y.F.

Peer review information Nature Communications thanks the 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) 2021