7094–7104 *Nucleic Acids Research, 2019, Vol. 47, No. 13* *Published online 27 May 2019*
doi: 10.1093/nar/gkz470

Structures and mechanism of transcription initiation by bacterial ECF factors

Chengli Fang[1,2,†], Lingting Li[1,2,†], Liqiang Shen[1,2], Jing Shi[3], Sheng Wang[4,*], Yu Feng[3,*] and Yu Zhang[1,*]

[1]Key Laboratory of Synthetic Biology, CAS Center for Excellence in Molecular Plant Sciences, Shanghai Institute of Plant Physiology and Ecology, Chinese Academy of Sciences, Shanghai 200032, China, [2]University of Chinese Academy of Sciences, Beijing 100049, China, [3]Department of Biochemistry and Molecular Biology, School of Medicine, Zhejiang University, Hangzhou 310058, China and [4]Computational Bioscience Research Center (CBRC), King Abdullah University of Science and Technology (KAUST) Thuwal, 23955, Saudi Arabia

Received February 08, 2019; Revised May 09, 2019; Editorial Decision May 14, 2019; Accepted May 17, 2019

ABSTRACT

Bacterial RNA polymerase (RNAP) forms distinct holoenzymes with extra-cytoplasmic function (ECF) σ factors to initiate specific gene expression programs. In this study, we report a cryo-EM structure at 4.0 Å of *Escherichia coli* transcription initiation complex comprising σ^E—the most-studied bacterial ECF σ factor (*Ec* σ^E-RPo), and a crystal structure at 3.1 Å of *Mycobacterium tuberculosis* transcription initiation complex with a chimeric $\sigma^{H/E}$ (*Mtb* $\sigma^{H/E}$-RPo). The structure of *Ec* σ^E-RPo reveals key interactions essential for assembly of *E. coli* σ^E-RNAP holoenzyme and for promoter recognition and unwinding by *E. coli* σ^E. Moreover, both structures show that the non-conserved linkers (σ_2/σ_4 linker) of the two ECF σ factors are inserted into the active-center cleft and exit through the RNA-exit channel. We performed secondary-structure prediction of 27,670 ECF σ factors and find that their non-conserved linkers probably reach into and exit from RNAP active-center cleft in a similar manner. Further biochemical results suggest that such σ_2/σ_4 linker plays an important role in RPo formation, abortive production and promoter escape during ECF σ factors-mediated transcription initiation.

INTRODUCTION

Bacterial σ factors are key components of the bacterial RNAP holoenzyme. During transcription initiation, the σ factors associate with RNAP core enzyme, guide the transcription machinery to promoter regions of genes, unwind double-strand promoter DNA, and facilitate *de novo* RNA synthesis (1–4). The genomes of bacteria comprise one primary σ factor (or group-1 σ factor; σ^{70} in *Escherichia coli*) maintaining expression of majority of genes, and a collection of alternative σ factors in control of subsets of genes responding to certain intracellular and environmental signals (5,6).

The group-1 σ factor is the most studied and well-known σ factors, *E. coli* σ^{70} is composed of multiple domains—$\sigma_{1.1}$, $\sigma_{1.2}$, σ_{NCR}, σ_2, $\sigma_{3.1}$, $\sigma_{3.2}$ and σ_4. Domains $\sigma_{1.2}$, σ_{NCR}, σ_2, $\sigma_{3.1}$ and σ_4 reside on the surface of RNAP core enzyme and are responsible for recognizing promoter DNA (7–14). Domain σ_2 also initiates unwinding of double-stranded promoter DNA to form a transcription bubble (10,11,13,15). Domain $\sigma_{3.2}$, a linker between $\sigma_{3.1}$ and σ_4, threads the RNAP and makes extensive interactions in the active-center cleft. The $\sigma_{3.2}$ linker serves as a mimic of RNA to pre-organize template single-strand DNA (ssDNA) of the transcription bubble into a helical conformation (10), facilitating base-pairing of initiating NTPs to template ssDNA (16,17); however, the RNA-mimic $\sigma_{3.2}$ linker would inevitably collide with the 5′-end of nascent RNA of length >4 nt, partially accounting for abortive production, transcription initiation pausing (18–20) and promoter escape (21,22).

The alternative σ factors contain three groups of σs belonging to the σ^{70} family (group-2, 3 and 4 σs) and one group of σs belonging to the σ^{54} family (1). The group-2 σ factors (σ^{38} in *E. coli*) contain all domains except domain $\sigma_{1.1}$ and recognize promoters very similar to those of group-1 σ factor (primary σ factor). The group-3 σ factors (σ^{32} or σ^{28} in *E. coli*) lack domain $\sigma_{1.1}$ and $\sigma_{1.2}$ and recognize promoters distinct from those of group-1 factor. The group-4 σ factors (also known as Extra-Cytoplasmic Function σ factors; ECF σ factors) only retain conserved domains

*To whom correspondence should be addressed. Tel: +86 21 54924351; Fax: +86 21 54924352; Email: yzhang@sippe.ac.cn
Correspondence may also be addressed to Yu Feng. Tel: +86 571 88208093; Fax: +86 571 88208094; Email: yufengjay@zju.edu.cn
Correspondence may also be addressed to Sheng Wang. Tel: +1 773 7422870; Fax: +1 773 7422870; Email: realbigws@gmail.com
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

σ_2 and σ_4. ECF σ factors are the most abundant, compact and divergent σ factors (1,3). They are important for stress adaption of most bacteria and are associated with virulence and drug resistance of pathogenic bacteria (6,23–26). ECF σ factors recognize promoters with stringent specificity and have been engineered to orthogonal transcriptional elements for constructing gene circuits (27–29).

Escherichia coli σ^E (σ^{24}) is an essential ECF σ factor. It maintains cell envelope integrity both under stress conditions (heat-shock, acid or oxidative stresses) and during normal growth (30); it also participates in biofilm formation and drug resistance of pathogenic *E. coli* (31,32). The activation of σ^E is induced by mis-folded proteins in periplasm under cell envelope stress, which triggers a cascade of protease cleavage resulting in release of σ^E into cytoplasm (33). The σ^E subsequently forms a holoenzyme with RNAP and directly upregulates expression of ∼100 protein-encoding genes that are involved in transport and assembly of outer membrane proteins and lipo-polysaccharide to relieve stress. It also indirectly downregulates expression of outer membrane proteins by activating transcription of their small regulatory RNAs–MicA, RybB and MicL to reduce protein load (34,35).

Escherichia coli σ^E contains two conserved domains (σ_2 and σ_4) and a non-conserved σ_2/σ_4 linker as other bacterial ECF σ factors. *Escherichia coli* σ^E recognizes promoters with consensus sequences at the −35 and −10 elements of 'GGAACTT' and 'GTC', respectively (36–38). Crystal structures of *E. coli* $\sigma^E{}_2$ complexed with the −10 element promoter DNA and of *E. coli* $\sigma^E{}_4$ complexed with the −35 element promoter DNA reveal protein-DNA interactions for sequence-specific promoter recognition by *E. coli* σ^E (12,39). However, no structural information is available for a RNAP complex comprising *E. coli* σ^E except a model of σ^E-RNAP holoenzyme predicting that $\sigma^E{}_2$ and $\sigma^E{}_4$ bind to RNAP as primary σ factors do (40). It is unknown how RNAP and *E. coli* σ^E are assembled to a functional σ^E-RNAP holoenzyme, how the σ^E-RNAP recognizes promoter DNA, and how σ^E-RNAP initiates transcription.

Recently Li *et al.* and Lin *et al.* have independently reported crystal structures of *Mycobacterium tuberculosis* transcription initiation complexes comprising ECF σ factors (σ^H or σ^L) (41,42). The structures together revealed interactions among ECF σ factors, RNAP core enzyme and promoter DNA, and surprisingly showed that the σ_2/σ_4 linkers of the two ECF σ factors interact with RNAP core enzyme in an analogous way as the $\sigma_{3.2}$ of primary σ factor does—the linker inserts into the active center cleft and exits out through the RNA exit channel (43,44). As the σ_2/σ_4 linkers of ECF σ factors are highly divergent in length and sequence, it is intriguing to know whether the σ_2/σ_4 linkers of other ECF σ factors interact with RNAP in a similar manner, how RNAP manages to accommodate the extremely variable structure modules using one binding site, and more importantly what role the linkers play during transcription initiation.

In this study, we determined a cryo-EM structure at 4.0 Å of *E. coli* transcription initiation complex comprising *E. coli* σ^E, and a crystal structure of *M. tuberculosis* transcription initiation complex comprising a chimeric *M. tuberculosis* $\sigma^{H/E}$ factor. The structures reveal protein-protein interactions essential for RNAP holoenzyme assembly, and protein-DNA interactions critical for promoter recognition and unwinding by *E. coli* σ^E. More importantly, the structures show that the σ_2/σ_4 linkers of *E. coli* σ^E and *M. tuberculosis* σ^E insert into the active-center cleft of RNAP and interact with template single-stranded DNA as do the σ_2/σ_4 linkers of *M. tuberculosis* σ^H and σ^L, despite no sequence similarity of the linker regions. The structure prediction of 27,670 bacterial ECF σ factors shows that the σ_2/σ_4 linkers of ECF σ factors retain similar secondary structures at the end regions, indicating that the σ_2/σ_4 linkers, albeit highly divergent in sequence, probably follow the same path to enter and exit the active center of RNAP. We demonstrated that the σ_2/σ_4 linker is essential for ECF σ-initiated transcription probably by facilitating several steps including RPo formation, synthesis of initial short RNA transcripts, and promoter escape.

MATERIALS AND METHODS

Plasmids

DNA fragments encoding *E. coli* σ^E, *Bacillus subtilis* σ^W, *B. subtilis* σ^M, *M. tuberculosis* σ^H, and *M. tuberculosis* σ^E were amplified from genomic DNA of *E. coli*, *B. subtilis* and *M. tuberculosis* respectively. The DNA fragments were subsequently cloned into pTolo-EX5 (ToloBio Inc.) or pET28a as described in Supplementary Table S1. The pEASY-p*rpoH*, pEASY-p*rpoE,* pEASY-p*sigM,* pEASY-p*sigW,* pEASY-p*sigB* and pEASY-p*ClpB* were constructed by inserting the promoter region (−50 to +50) of respective genes amplified from genomic DNA into the pEASY-blunt vector (Transgen biotech, China).

Proteins

The wild-type or derivatives of bacterial ECF σ factors were over-expressed in *E. coli* BL21(DE3) cells (NovoProtein), and purified from soluble fractions using Ni-NTA (SMART, Inc.) and Heparin columns (GE Healthcare). The *Mtb* $\sigma^E{}_2$ and *Bs* $\sigma^E{}_2$ were obtained from the inclusion body. The *E. coli*, *M. tuberculosis*, and *B. subtilis* RNAP core enzymes were over-expressed in *E. coli* BL21(DE3) and sequentially purified on a Ni-NTA affinity column, a Mono Q ion-exchange column, and a Superdex S200 size-exclusion column.

Crystallization and structure determination of *Mtb* σ^H–RPo and $\sigma^{H/E}$–RPo

The *Mtb* σ^H–RPo and $\sigma^{H/E}$–RPo complexes for crystallization were prepared by reconstitution. The *Mtb* RNAP core enzyme, *Mtb* σ^H(or $\sigma^{H/E}$), and nucleic-acid scaffolds (Figure 3A) were mixed at 1: 4: 1.2 molar ratio and incubate at 4°C overnight. The RPo complexes were purified using a Hiload 16/60 Superdex S200 column (GE Healthcare, Inc.) and stored in 20 mM Tris–HCl pH 8.0, 0.1 M NaCl, 1% (v/v) glycerol, 1 mM 1,4-dithiothreitol (DTT) with a concentration of 7.5 mg/ml. Crystals of *Mtb* σ^H–RPo were obtained from 0.08 M Magnesium acetate, 0.05

M sodium cacodylate pH 6.5, 15% PEG400; and crystals of *Mtb* $\sigma^{H/E}$–RPo were obtained from 0.2 M sodium acetate, 0.1 M sodium citrate pH 5.5, 10% PEG4000. The X-ray diffraction data were collected at Shanghai Synchrotron Radiation Facility (SSRF) beamlines 17U and 19U, and the structures were solved by molecular replacement with Phaser MR using the structure of *M. tubercolusis* RNAP holo enzyme (PDB ID: 5ZX3).

Cryo-EM structure determination of *E. coli* σ^E-RPo

The *E. coli* σ^E-RPo were obtained by reconstitution with *E. coli* RNAP core enzyme, *E. coli* σ^E, and nucleic-acid scaffold as above (Figure 1A). The *E. coli* σ^E-RPo were concentrated to ∼15 mg/ml and stored in 10 mM Hepes pH 7.5, 50 mM KCl, 5 mM $MgCl_2$, 3 mM DTT. The *E. coli* σ^E-RPo was mixed with CHAPSO (Hampton Research Inc.) to a final concentration of 8 mM prior to grid preparation. The complex (3 μl) were subsequently applied on a glow-discharged C-flat CF-1.2/1.3 400 mesh holey carbon grids (Protochips, Inc.), and plunge-frozen in liquid ethane using a Vitrobot Mark IV (FEI). The grids were loaded into a 300 keV Titan Krios (FEI) equipped with a K2 Summit direct electron detector (Gatan) and a dataset was collected. The electron density map was obtained by single-particle reconstitution with RELION2.1. Gold-standard Fourier-shell-correlation analysis indicated a mean map resolution of 4.02 Å. The structure model was built in Coot and refined in Phenix.

Stopped-flow assay

The promoter for the stop flow assay was prepared as in Supplementary Figure S5A. To monitor the efficiency of RPo formation of *E. coli* RNAP holoenzymes comprising wild-type or derivatives of *E. coli* σ^E, 60 μl σ^E-RNAP holoenzyme (200 nM) and 60 μL Cy3-P*rpoE* (4 nM) in 10 mM Tris–HCl, pH 7.7, 20 mM NaCl, 10 mM MgCl2, 1 mM DTT were rapidly mixed and the change of Cy3 fluorescence was monitored in real time by a stopped-flow instrument (SX20, Applied Photophysics Ltd, UK) equipped with a excitation filter (515/9.3 nm) of and a long-pass emission filter (570 nm). The data were plotted in SigmaPlot (Systat software, Inc.) and the observed rates $Kobs,_1$ and $Kobs,_2$ were estimated as in the Supplementary Materials and Methods.

Fluorescence polarization (FP) competitive assay

The *E. coli* σ^E was labeled with fluorescein at residues C165. The affinity of *E. coli* RNAP core enzyme and wild-type *E. coli* σ^E was first determined as ∼53 nM by a FP assay. A FP competition assay was further employed to compare the affinities of wild-type and derivatives of *E. coli* σ^E to RNAP core enzyme. Label-free *E. coli* σ^E (0, 2.5, 5, 10, 20, 40, 80, 160, 320, 640, 1280, 2560 or 5120 nM; final concentration) pre-mixed with fluorescein-labeled *E. coli* σ^E (5 nM; final concentration) were incubated with *E. coli* RNAP core enzyme (100 nM; final concentration) in FP buffer at room temperature for 20 min in FP buffer. The FP signals were measured using a plate reader (SPARK, TECAN Inc.) equipped with excitation filter of 495/10 nm and emission filter of 520/20. The data were plotted in SigmaPlot (Systat software, Inc.) and the IC_{50} were estimated as in the Supplementary Materials and Methods.

RESULTS

The cryo-EM structure of *E. coli* σ^E-RPo

To obtain a structure of *E. coli* σ^E-RPo, we reconstituted the *E. coli* σ^E-RPo complex with *E. coli* RNAP core enzyme, *E. coli* σ^E and a nucleic-acid scaffold (Figure 1A and B; Supplementary Figure S1A). The nucleic-acid scaffold (−34 to +14; with respect to +1 as transcription start site) is composed of a 24-bp upstream double-stranded DNA (dsDNA) with consensus sequences of the −35 element, a 12-bp transcription bubble (maintained open by having non-complimentary sequences on nontemplate- and template-strand DNA), a 12-bp downstream dsDNA, and a 5-mer RNA.

We obtained a cryo-EM map at 4.0 Å for the *E. coli* σ^E-RPo complex with local resolution at the active-center cleft of RNAP around ∼3 Å (Figure 1C; Supplementary Figure S1B–E and Table S3). The map shows clear density for residues of $\sigma^E{}_2$ (residues 5–87) and $\sigma^E{}_4$ (residues 131–190) and all residues of the $\sigma^E{}_2$/ $\sigma^E{}_4$ linker (residues 88–130) (Figure 1E and Supplementary Figure S2B). The map also shows clear density for the upstream dsDNA, template and nontemplate ssDNA of transcription bubble, the RNA/DNA hybrid and the downstream dsDNA (Figure 1F and Supplementary Figure S2D). The RNAP clamp in the structure of σ^E-RPo adopts a closed conformation as other bacterial transcription RPo complexes (Supplementary Figure S3A) (10,45–48). The template ssDNA and nontemplate ssDNA follow the same path as in the structure of *M. tuberculosis* σ^H-RPo (Supplementary Figure S3B) (41).

In the structure of σ^E-RPo, the domains $\sigma^E{}_2$ and $\sigma^E{}_4$ locate on the surface of RNAP (Figure 1D). $\sigma^E{}_2$ attaches to the clamp helices of the RNAP-β' subunit *via* a polar surface as the σ_2 of other σ^{70}-family factors does (Figure 1D and Supplementary Figure S3C) (48,49). The residues on the interface are conserved (Supplementary Figure S4A). Intriguingly, $\sigma^E{}_4$ uses a distinct hydrophobic surface to bind the tip helix of RNAP-β flap domain (βFTH). The interface residues include V131, F132, I135, L151, I181, V185 and I189 of $\sigma^E{}_4$ and E898, L901, L902, I905 and F906 of the βFTH. Such interaction induces a 90° rotation of the βFTH, where the βFTH is further stabilized by the extended hydrophobic surface created by residues I121, L123 and L127 of the $\sigma^E{}_2/\sigma^E{}_4$ linker (Figure 2A and B).

The promoter recognition and unwinding by *E. coli* σ^E

The structure of *E. coli* σ^E-RPo is superimposable on the binary structure of *E. coli* $\sigma^E{}_4$/−35 element promoter dsDNA (Figure 2C), supporting the previous conclusion that σ_4 of bacterial ECF σ factors reads sequence and shape of −35 dsDNA (12). The structure of *E. coli* σ^E-RPo is also superimposable on the binary structure of *E. coli* $\sigma^E{}_2$/−10 element promoter ssDNA (Figure 2D). In particular, the T_{-10} and C_{-9} of the non-template strand were inserted into two protein pockets (Figure 2E) in the exactly same manner

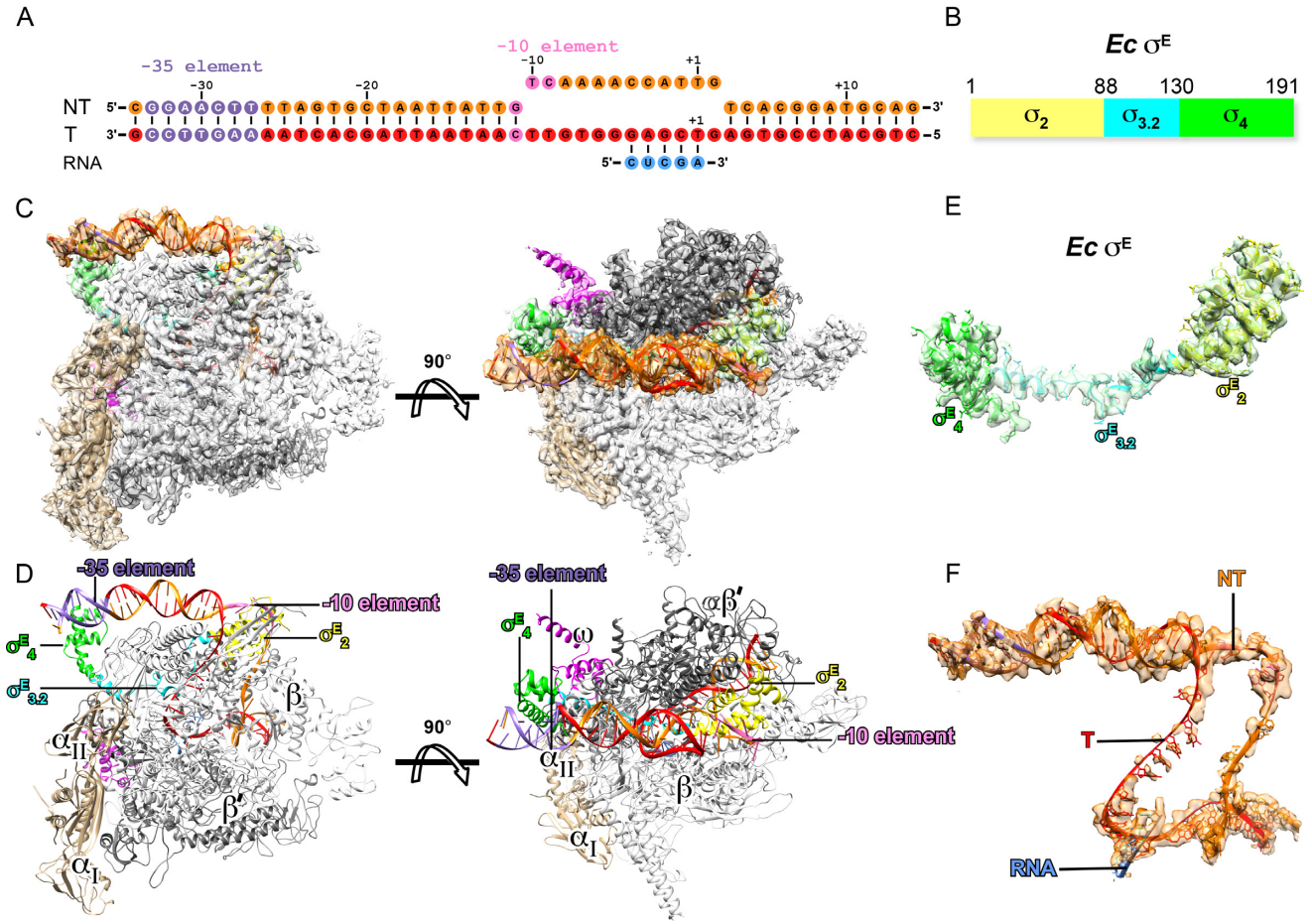

Figure 1. The cryo-EM structure of *Escherichia coli* σ^E-RPo. (**A**) The nucleic-acid scaffold used for structure determination. (**B**) Schematic diagram of *E. coli* σ^E. (**C**) The side and top views of cryo-EM electron density map (gray surface) at resolution 4.0 Å and the structure model of σ^E-RPo. (**D**) The side and top views of the overall structure of *E. coli* σ^E-RPo. (**E**) The electron density map and structural model of *E. coli* σ^E. (**F**) The electron density map and structural model of the nucleic-acid scaffold. RNAP-α subunit, light brown; RNAP-β subunit, gray; RNAP-β' subunit, dark gray; RNAP-ω subunit, pink; σ^E_2, light green; the $\sigma_{3.2}$-like linker of σ^E ($\sigma^E_{3.2}$), cyan; σ^E_4, green. Template DNA, red; non-template DNA, orange; RNA, blue; the −35 element DNA, light blue; the −10 element, violet.

as in the structure of *E. coli* σ^E_2/−10 ssDNA. The DNA–protein interactions are sequence specific, as swapping the 'specificity loop' of *E. coli* σ^E altered the specificity of the element (39).

The structure implicates that N80 might serve as a wedge to separate the base pair at position −10. To explore contributions of the residue to promoter unwinding, we modified a stopped-flow assay to monitor the RPo formation by *E. coli* σ^E-RNAP, in which the fluorescence of a Cy3 fluorophore at +1 position on non-template strand DNA increases upon RPo formation (Figure 2F and Supplementary Figure S5A). Similar assays have been used to measure the kinetics of RPo formation by the primary σ factor (50–52). As shown in Figure 2F, the fluorescence rapidly increases and reaches to a plateau in 5 seconds after mixing the σ^E-RNAP with promoter DNA, while RNAP core enzyme induces no change of fluorescence, validating the assay. The kinetics of RPo equilibration is two times slower by σ^E(N80A)-RNAP holoenzyme compared with wild-type σ^E-RNAP, suggesting a role of N80 during RPo formation probably by facilitating promoter unwinding (Figure 2F). Interestingly, mutations of the protein pockets on σ^E for T_{-10} and C_{-9} (F64A or W73A) also exhibited slowed RPo equilibration (Figure 2F), indicating that the RPo equilibration could be accelerated by securing the unwound nucleotides. It is worth noting that all curves could be perfectly fitted with a typical two-phase kinetics (a fast phase and a slow phase), suggesting the existence of a significant intermediate (RPi) on the path toward RPo (Figure 2F and Supplementary Figure S5C–E). Alanine substitutions of N80, F64 or W73 slow down kinetics of both phases (Supplementary Table S5).

The above evidence supports the conclusion that *E. coli* σ^E unwinds promoter at the −11/−10 junction in a previous study (39), similar to *M. tuberculosis* σ^H (41), but different from *E. coli* σ^{70} (13), which unwinds promoter DNA at a position 1-bp downstream of that by the ECF σ factors (Supplementary Figure S3E–H) (9,41,42). Structure superimposition (*M. tuberculosis* σ^H-RPo, *E. coli* σ^{70}-RPo, and *E. coli* σ^E-RPo) reveals that the melting residues of the primary of σ^A (W433 and W434 for *E. coli* σ^{70}) and ECF σ factors (N80 for *E. coli* σ^E) locate at slightly different po-

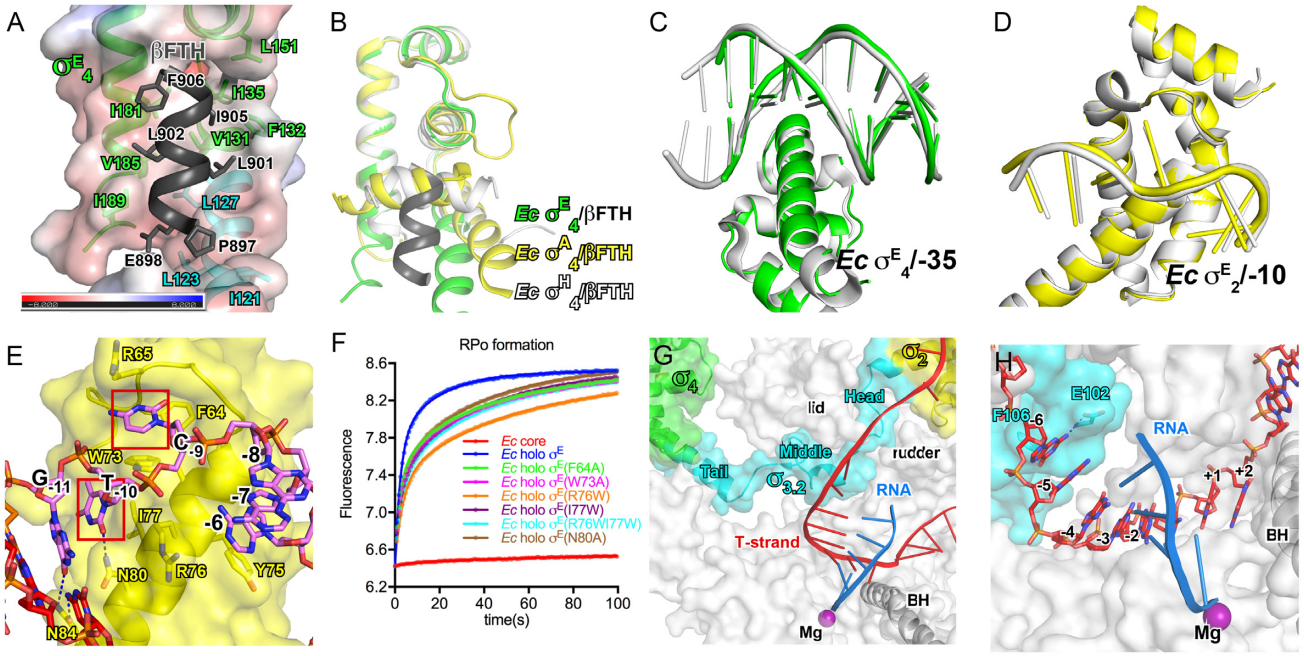

Figure 2. The interactions among RNAP core enzyme, σ^E and nucleic-acid scaffold. (**A**) The βFTH interacts with a large hydrophobic surface created by σ^E_4 and the $\sigma_{3.2}$-like linker of σ^E. (**B**) The structural comparison of interactions between βFTH and domain σ_4 of *Escherichia coli* σ^E-RPo (green and dark gray), *E. coli* σ^A-RNAP (PDB ID: 6CA0; yellow) and *Mycobacterium tuberculosis* σ^H-RPo (PDB ID: 5ZX2; gray). (**C**) The *E. coli* σ^E–RPo (green) is superimposable to the crystal structure of *E. coli* σ^E_4/-35 dsDNA (gray). (**D**) The *E. coli* σ^E-RPo (yellow) is superimposable to the crystal structure of *E. coli* σ^E_2/-10 ssDNA (gray); (**E**) The detailed interaction between the *E. coli* σ^E_2 and the −10 element promoter DNA. (**F**) The stopped-flow experiments measuring the kinetics of promoter unwinding by WT or derivatives of *E. coli* σ^E-RNAP. The data points were recorded every 0.1 s and the data were fitted as described in 'Materials and Methods' section. The experiments were repeated for three times and representative curves are shown. (**G**) The $\sigma_{3.2}$-like linker of σ^E ($\sigma^E_{3.2}$) inserts into the active-center cleft. (**H**) The detailed interactions between $\sigma^E_{3.2}$ and template ssDNA of the transcription bubble.

sitions on the protein surface (Supplementary Figure S3E–H). Tryptophan substitution of the residues of *E. coli* σ^E locating at the corresponding positions of the W-dyad on σ^A (R76W, I77W or R76W/I77W) resulted in substantial decrease of promoter unwinding efficiency, confirming that σ^E opens promoter through a different mechanism than primary σ factor (Figure 2F and Supplementary Table S5).

The σ^E_2/σ^E_4 linker interacts with the active-center cleft of RNAP

We discovered that the σ^E_2/σ^E_4 linker dives into the active-center cleft of RNAP and emerges out through the RNA-exit channel (Figure 2G). The path inside of RNAP of the σ^E_2/σ^E_4 linker is remarkably similar to that of the $\sigma_{3.2}$ of the group-1 σ factor and also to those of linker regions of two other ECF σ factors (*M. tuberculosis* σ^H and σ^L) (41,42), therefore, we designated the σ^E_2/σ^E_4 linker as $\sigma^E_{3.2}$-like linker (Figure 2G and Supplementary Figure S6). The $\sigma^E_{3.2}$-like linker region could be further divided into three sub-regions—the head (residues 88–98), middle (residues 99–118) and tail (residues 119–130) (Figure 2G). The head sub-region extends the helix of σ^E_2 and enters the active-center cleft through T-ss DNA channel created by the RNAP-β' lid and rudder motifs. The middle sub-region passes underneath the lid domain and makes a turn toward the RNA-exit channel; it resides in the RNAP active-center cleft and contacts the T-6 nucleotide (Figure 2H). The tail sub-region forms a continuous helix with the first helix of σ^E_4 and exits the RNAP active-center cleft through the RNA-exit channel (Figure 2G).

The $\sigma_{3.2}$-like linker of *M. tuberculosis* σ^E also inserts into the active-center cleft of RNAP

Considering the fact that there is no similarity on primary sequences of the σ_2/σ_4 linker regions of bacterial ECF σ factors (1), we are interested to know whether the linker regions of other bacterial ECF σ factors follow the same path in RNAP. The initial attempts to obtain additional structures of RNAP complexed with ECF σ factors failed. Inspired by the results that chimeric σ factors with the linker region swapped function normally and the idea of determination of crystal structures of transcription initiation complexes containing chimeric σ factors (41,42), we sought to obtain crystal structures of the *Mtb* RPo complexes with chimeric σ^H factors. We first took advantage of the high crystallizability of *M. tuberculosis* σ^H–RPo and obtained a novel robust crystal form of *M. tuberculosis* σ^H–RPo (*Mtb* σ^H-RPo-Fork) at 2.9 Å (Supplementary Table S2) (41). The *Mtb* σ^H-RPo is reconstituted *in vitro* with *M. tuberculosis* RNAP σ^H-holoenzyme, a downstream fork DNA scaffold and a 5-mer RNA (Figure 3A). The *Mtb* σ^H in the structure of *Mtb* σ^H-RPo-Fork makes the same interactions with RNAP core enzyme and with promoter DNA as it does in the previously reported structure of *Mtb* σ^H-RPo with a full transcription bubble promoter DNA (41).

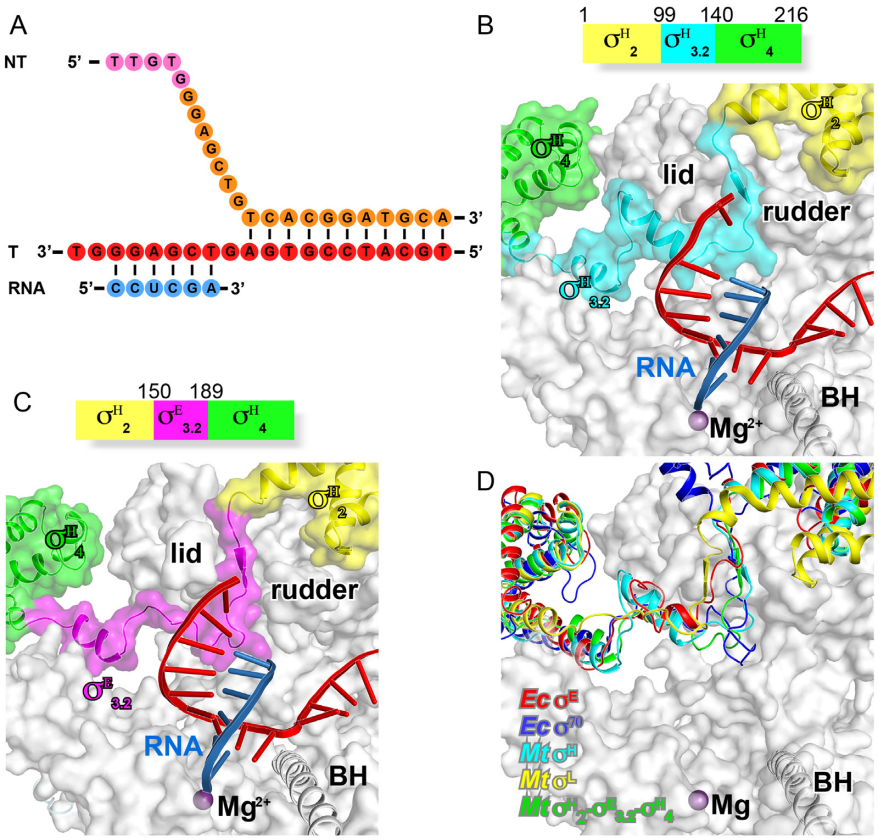

Figure 3. The crystal structure of *Mycobacterium tuberculosis* σ^{H}-RPo and $\sigma^{H/E}$-RPo. (**A**) The nucleic-acid scaffold used for the structure determination. (**B**) The schematic diagram of *M. tuberculosis* σ^{H} and interaction of the $\sigma_{3.2}$-like linker of σ^{H} ($\sigma^{H}_{3.2}$) with RNAP active-center cleft. (**C**) The schematic diagram of *M. tuberculosis* $\sigma^{H/E}$ and interaction of $\sigma^{E}_{3.2}$ with RNAP active-center cleft. (**D**) The domain $\sigma_{3.2}$ of *Ec* σ^{E}, *Ec* σ^{70}, *Mtb* σ^{E}, *Mtb* σ^{L}, *Mtb* σ^{H} factors follow similar path to enter and exit RNAP active-center cleft.

By using the same fork scaffold, we determined a crystal structure at 3.1 Å of *Mtb* $\sigma^{H/E}$-RPo comprising the same nucleic-acid scaffold and a chimeric $\sigma^{H/E}$ with σ_2/σ_4 linker of σ^{H} replaced by that of *Mtb* σ^{E} (Figure 3C). In the structure of *Mtb* $\sigma^{H/E}$-RPo, the σ_2/σ_4 linker region of *Mtb* σ^{E} follows a similar path through RNAP active-center cleft and makes interactions with the template ssDNA as other bacterial ECF σ factors, providing another evidence for the conserved interaction mode of the linker region with RNAP (Figure 3D).

The head and tail of $\sigma_{3.2}$-like linkers retain conserved secondary structures

All of the four available structures comprising bacterial ECF σ factors (*Ec* σ^{E}-RPo and *Mtb* $\sigma^{H/E}$-RPo in this study and *Mtb* σ^{L}-RPo and *Mtb* σ^{H}-RPo in previous reports) show that $\sigma_{3.2}$-like regions of ECF σ factors interact with RNAP in a similar manner (Supplementary Figure S6A–D), although the four σ factors share little similarity of primary sequences in the regions. Multiple sequence alignments (MSAs) of the 27,670 ECF σ factors reveal a clear boundary of their $\sigma_{3.2}$-like region (residues 88–130 for *E. coli* σ^{E}) and confirmed that the linker is the least conserved region of ECF σ factors (Figure 4A and Supplementary Figure S4; Supplementary Files 1 and 2). However, structural comparison of the $\sigma_{3.2}$-like linkers of the four available RPo structures comprising ECF σ factors exhibits similar secondary structures for the head and tail sub-regions. Namely, the head sub-regions contain a short helix followed by a short β strand or a coil; while the tail sub-regions are mainly composed of a helix (Figure 4B and C).

To explore whether other bacterial ECF σ factors also retain similar secondary structure folds for the $\sigma_{3.2}$-like linker regions. We performed secondary-structure prediction of the $\sigma_{3.2}$-like linker regions of the 27,670 bacterial ECF σ factors using RaptorX-Property and calculated the probability score of secondary structures for each position (53,54). The predictions agree very well with the secondary-structure pattern of the four available structures (Supplementary Figure S7); 85% of residues adopt exactly the same secondary structures as predicted, validating the predictions. More importantly, the predictions show a strikingly conserved pattern of secondary structures for the head and tail sub-regions of $\sigma_{3.2}$-like linkers. Namely, ∼80% of ECF σ factors are predicted to contain a short helix followed by a coil in the head sub-region and a short helix in the tail sub-region of $\sigma_{3.2}$-like linkers (Figure 4D and Supplementary File 3).

The conserved helical structures of the head and tail sub-regions of $\sigma_{3.2}$-like linkers strongly implicate that $\sigma_{3.2}$-like linkers of most bacterial ECF σ factors probably bind to

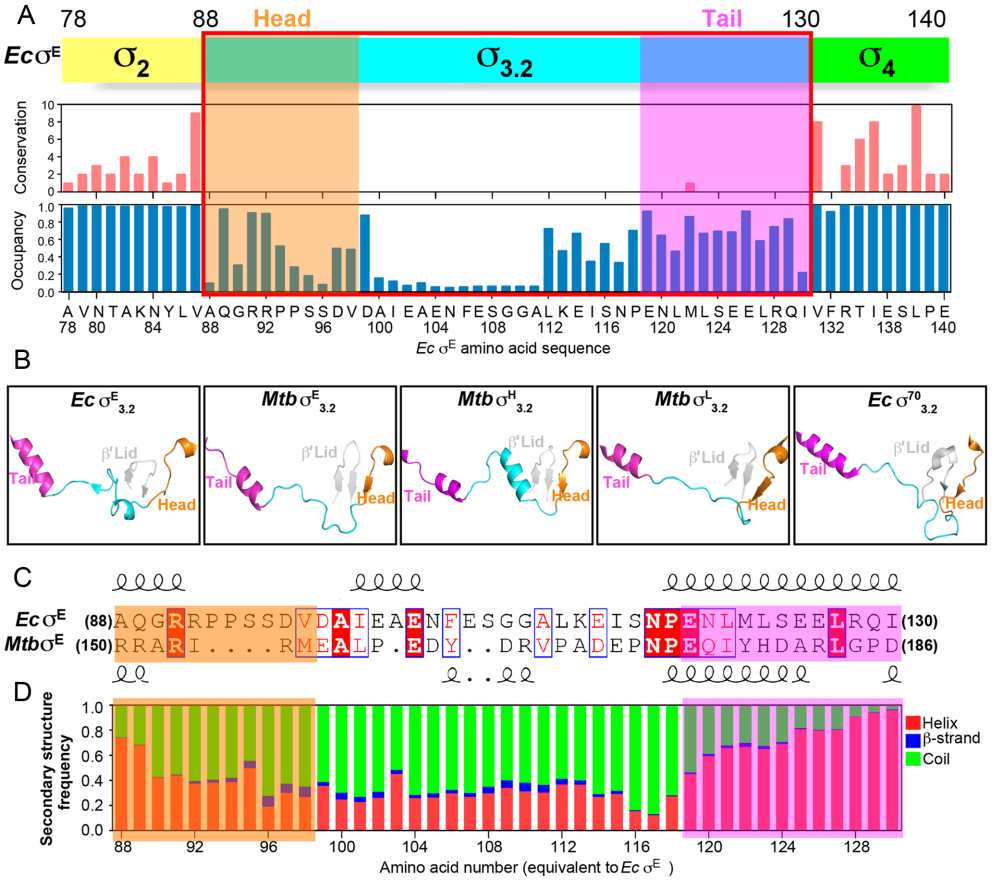

Figure 4. The sequence alignment and secondary-structure prediction of 27,670 bacterial ECF σ factors. (**A**) The sketched diagram of the MSA of 27,670 bacterial ECF σ factors. Upper panel, the schematic diagram of *Escherichia coli* σ^E; middle panel, the conservation score from the MSA for each position of *E. coli* σ^E; bottom panel, the occupancy score from MSA for each position of *E. coli* σ^E. The conservation and occupancy scores were calculated by Jalview. The occupancy scores show the ratio of ungapped positions in each column of the alignment. (**B**) The structures of the head, middle and tail sub-regions of *Ec* $\sigma^E{}_{3.2}$, *Mtb* $\sigma^E{}_{3.2}$, *Mtb* $\sigma^H{}_{3.2}$ (PDB ID: 5ZX3) and *Mtb* $\sigma^L{}_{3.2}$ (PDB ID: 6DV9), and *Ec* $\sigma^{70}{}_{3.2}$(PDB ID: 6CA0). (**C**) The primary protein sequences and secondary structures of *E. coli* $\sigma^E{}_{3.2}$ and *Mtb* $\sigma^E{}_{3.2}$. (**D**) The probability score of secondary structures of $\sigma_{3.2}$-like linkers of bacterial ECF σ factors at corresponding positions of $\sigma^E{}_{3.2}$.

the RNAP active-center cleft. The head sub-region of $\sigma_{3.2}$-like linkers extends the last helix of domain σ_2, and precisely guides the rest of $\sigma_{3.2}$-like linkers into the RNAP active-center cleft through T-ssDNA channel as those of *M. tuberculosis* σ^H, σ^E, σ^L and *E. coli* σ^E (Figures 2G, 3B and C); while the short coil of the head sub-region passes through the T-ssDNA channel probably by forming a β-sheet with the RNAP β'-lid motif as those of *M. tuberculosis* σ^H, σ^L and σ^E (Figure 4B). Similarly, the first helix of σ_4 together the tail helix of $\sigma_{3.2}$-like linkers reaches into the RNA exit channel, where it connects with the middle sub-region of $\sigma_{3.2}$-like linkers as those of *M. tuberculosis* σ^H, σ^L, σ^E and *E. coli* σ^E (Figures 2G, 3B-C and 4B).

The $\sigma_{3.2}$-like linker plays pivotal role during transcription initiation

Above evidence suggests that $\sigma_{3.2}$-like linker of most bacterial ECF σ factors probably follows a similar path to enter into and exit from the active-center cleft of RNAP, implicating that such $\sigma_{3.2}$-like linker is an indispensable domain and probably plays essential function. We next explored the functional importance of the $\sigma_{3.2}$-like linker. We prepared wild-type and derivatives of well-studied bacterial ECF σ factors (including *E. coli* σ^E, *B. subtilis* σ^W, *B. subtilis* σ^M, *M. tuberculosis* σ^E and *M. tuberculosis* σ^H) and performed *in vitro* transcription experiments. The results of Figure 5A clearly showed that deleting or replacing the $\sigma_{3.2}$-like linker with a disordered sequence completely abolished the transcription activity of all tested bacterial ECF σ factors. The results suggest that the $\sigma_{3.2}$-like linker region is indeed essential for the transcription activity of bacterial ECF σ factors.

To further dissect the steps $\sigma_{3.2}$-like linker might be involved in during transcription initiation, we studied the assembly of RNAP holoenzyme, the formation of RPo, and the synthesis of abortive and productive transcripts by using wild-type or derivatives of *E. coli* σ^E. We developed a competitive FP assay (in which the unlabeled wild-type or derivatives of σ^E compete with [C165-FAM]σ^E for binding to RNAP core enzyme) to compare binding affinities of various σ factors. The *E. coli* σ^A exhibited the strongest inhibition with an IC_{50} ~5-fold lower than *E. coli* σ^E, which is in consistent with the previous finding that σ^A has higher affinity than that of other ECF σ factors (red in Figure

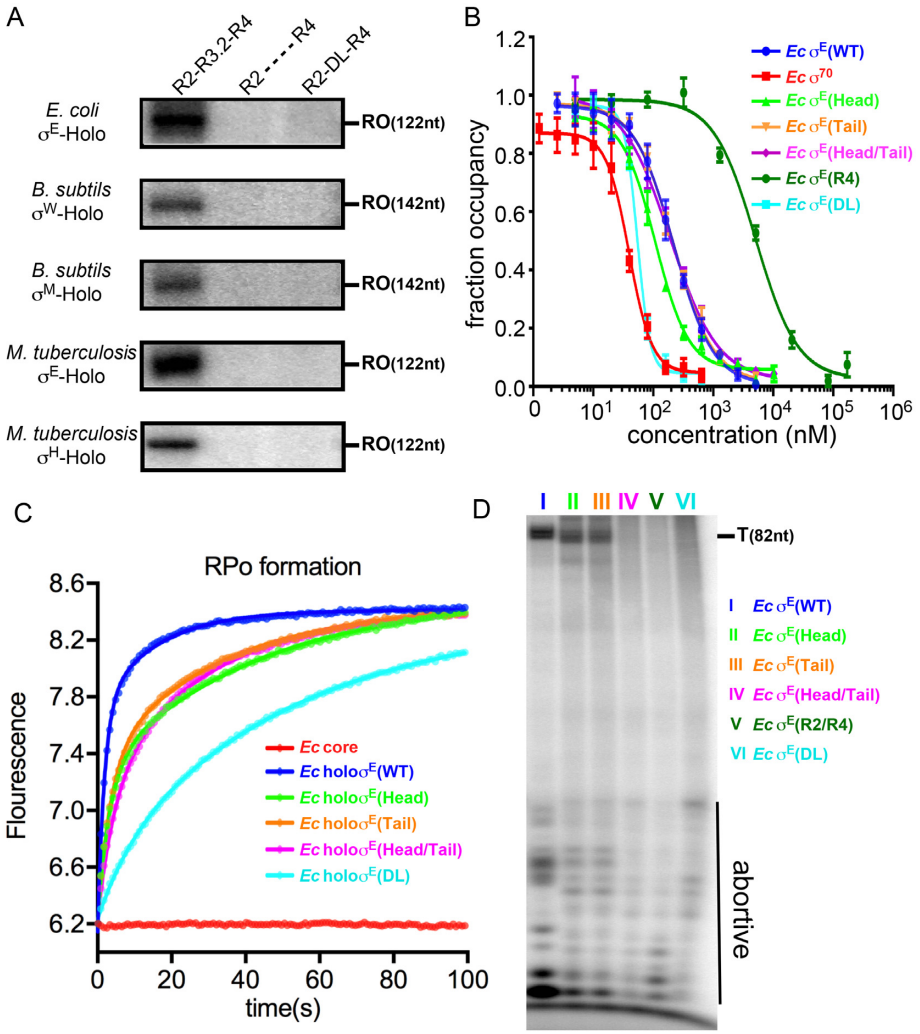

Figure 5. The $\sigma_{3.2}$-like linker plays essential roles in multiple steps of transcription initiation. (**A**) The *in vitro* transcription assay showing the transcription activity of WT and derivatives of various bacterial ECF σ factors. RO represents the run-off transcripts. (**B**) The FP competitive assay showing binding affinities of WT and derivatives of *Escherichia coli* σ^E to bacterial RNAP core enzyme. The experiments were repeated in triplicate, and the data are presented as mean ± S.E.M. (**C**) The stopped-flow experiments measuring the kinetics of promoter unwinding by WT or derivatives of *E. coli* σ^E. The data points were recorded every 0.1 s and the data were fitted as described in 'Materials and Methods' section. The *Ec* σ^E head region (residues 88–98) was replaced by 'GGSSGSGGGSSS' resulting in *Ec* σ^E (head); the $\sigma^E_{3.2}$ tail region (residues 119–130) was replaced by 'GGSSGSGGGSSS' resulting in *Ec* σ^E (tail); *E* $\sigma^E_{3.2}$ head region (residues 88–98) and tail region (residues 119–130) were replaced by 'GGSSGSGGSSS' and 'GGSSGSGGGSSS', respectively resulting in *Ec* σ^E (head/tail). (**D**) The *in vitro* transcription assay with WT or derivatives of *E. coli* σ^E. The 'abortive' represents abortive transcripts and the 'T' represents terminated transcripts of 82 nt. The *in vitro* transcription and stopped-flow experiments were repeated for three times and representative data are shown. The FP competitive experiments were repeated for three times and the data were presented as mean ± S.E.M.

5B and Supplementary Table S4) (55). Deletion of the $\sigma_{3.2}$-like linker of σ^E substantially decreases the binding affinity with an IC_{50} ~20-fold higher than *E. coli* σ^E (Figure 5B and Supplementary Table S4). However, replacing the head or the tail sub-regions of the $\sigma_{3.2}$-like linker of σ^E with random sequences has no significant change on the affinity of *E. coli* σ^E; while replacing the entire linker with a disordered acidic loop instead slightly increased the binding affinity. The results suggest that the presence of a physical linker—regardless of protein sequences of the linker—between σ_2 and σ_4 is necessary for maintaining the high affinity of *E. coli* σ^E to RNAP core enzyme (the linker physically ties the σ^E_2 and σ^E_4 together and thus greatly increase the affinity of the two domains to RNAP), but the interactions of the linker with RNAP plays little role for assembly of RNAP holoenzyme. The results are also consistent with the fact that bacterial ECF σ factors show highest conservation scores for RNAP-contacting residues on σ_2 and σ_4, but show no conservation on any residues on $\sigma_{3.2}$-like linkers (Figure 4A and Supplementary Figure S4). The results also explain that the identities of the −10 element are exclusively recognized at the non-template strand of promoter DNA (10,41,42).

The chimeric *E. coli* σ^E factors serve as good materials for subsequent experiments, as they showed similar affinity to wild-type σ^E with RNAP core enzyme. Therefore, any effects can be attributed the altered conformation of the $\sigma_{3.2}$-like linker or interactions between the linker and RNAP. We

next studied the potential effect on RPo formation using the chimeric *E. coli* σ^E factors by a stopped-flow fluorescence assay as described above. All the chimeric *E. coli* σ^E factors showed slowed RPo equilibration (Figure 5C and Supplementary Table S6), suggesting a role of the $\sigma_{3.2}$-like linker during RPo formation.

To explore the potential role of the $\sigma_{3.2}$-like linker of σ^E on the steps following RPo formation, we performed *in vitro* transcription assays. As shown in Figure 5D, RNAP holoenzymes comprising chimeric *E. coli* σ^E factors produce substantially less amount of abortive as well as full-length products. Intriguingly, RNAP holoenzyme with σ^E (DL) (the whole linker replaced by a disordered loop), σ^E (Head/Tail) (the head and tail regions of the $\sigma_{3.2}$-like linker are replaced by disordered loops), or σ^E (R2/R4) (disconnected $\sigma^E{}_2$ and $\sigma^E{}_4$; the $\sigma_{3.2}$-like linker is completely truncated) still produced abortive transcripts, albeit less efficiently, but produced no full-length products (Lane IV, V and VI in Figure 5D), suggesting that the $\sigma_{3.2}$-like linker probably also affect the later step of transcription initiation (i.e. promoter escape).

DISCUSSION

In this work, we have solved a cryo-EM structure of *E. coli* σ^E-RPo at 4.0 Å, a crystal structure of *M. tuberculosis* σ^H–RPo at 2.9 Å, and a crystal structure of *M. tuberculosis* $\sigma^{H/E}$-RPo at 3.1 Å. We included a 5-nt RNA primer (complimentary to nucleotides of template ssDNA at positions −4 to +1) to stabilize the complexes, a strategy has been used previously for determination of bacterial RPo complexes (10,13,56). The conformation of the 5-bp hybrid in our structures is indistinguishable to that of the *bona fide* bacterial transcription initiation complexes with 5-nt RNA (16,48), although it is not an on-pathway state of transcription initiation.

The structure of *E. coli* σ^E-RPo reveals protein-protein interactions essential for σ^E-RNAP holoenzyme assembly, and protein-DNA interactions essential for promoter recognition and unwinding. More importantly, the four structures of transcription initiation complexes comprising ECF σ factors and secondary-structure prediction of available 27,670 ECF σ factors show that the $\sigma_{3.2}$-like linkers of most bacterial ECF σ factors retain conserved pattern of secondary structures of the head and tail sub-regions and strongly suggest the $\sigma_{3.2}$-like linkers follow the same path to get in and out the active-center cleft of RNAP.

Our study explains how bacterial RNAP manages to accommodate such divergent $\sigma_{3.2}$-like linkers and why the primary sequences of $\sigma_{3.2}$-like linkers become so divergent during evolution. The head sub-region of $\sigma_{3.2}$-like linkers comprises a short helix followed by a coil. The short helix extends the last helix of σ_2 and help guide $\sigma_{3.2}$-like linker approaching into the channel to enter the active-center cleft of RNAP. The short coil forms a β-sheet with the lid domain of RNAP-β' subunit in three of four available structures of ECF σ-RPo (Figure 4B). Such interaction model explains the poor conservation of primary sequence in this region; as a β-sheet is typically stabilized through main-chain interactions. The tail sub-region of $\sigma_{3.2}$-like linkers in the RNA exit channel forms a long intact helix (occasionally with a kink) with residues of σ_4 (Figure 4B). It seems that the channels for entry and exit of $\sigma_{3.2}$-like linkers of ECF σ factors put some evolutionary pressure on the head or tail sub-regions and consequently certain secondary-structure patterns in the two sub-regions are retained. The middle sub-region of $\sigma_{3.2}$-like linkers locates mainly in the active-center cleft—a wide channel for accommodating DNA/RNA hybrid which puts much less restraint for indels on this sub-region during evolution, and thereby exhibits varied lengths in primary sequence and diverse secondary structures.

In case of primary σ factors, the $\sigma_{3.2}$ plays an essential role during transcription initiation (10,16,17,21,44,57). It inserts into the active-center cleft of RNAP, where it mimics an RNA molecule, pre-organizes the template ssDNA into a helical conformation, and increases the binding affinity of initiating NTPs. After showing that the $\sigma_{3.2}$-like linkers of bacterial ECF σ factors bind to the active-center cleft of RNAP and to the template ssDNA in the transcription bubble in a similar manner to the $\sigma_{3.2}$ of primary σ factors (Figures 2G and 3B-C), we demonstrated the $\sigma_{3.2}$-like linker of bacterial ECF σ factors is also crucial to transcription initiation as the $\sigma_{3.2}$ of primary σ factors. Deletion of the $\sigma_{3.2}$-like linker of bacterial ECF σ factors completely abolished production of full-length transcripts (Figure 5A). We further showed that multiple steps of transcription initiation require proper engagement of the $\sigma_{3.2}$-like linker in the active-center cleft of RNAP, as disrupting such interactions resulted in impaired ability to form RPo complex, synthesis of abortive transcripts as well as promoter escape (Figure 5).

Transcription machineries from all three domains of lives retain similar essential structure modules as the domain $\sigma_{3.2}$ of σ^{70} family (42)—domain RII.3 of σ^{54} family in bacteria (58,59), B-reader of TFIIB in archaea RNAP (60), reader of Rrn7 for eukaryotic Pol I (61,62), the B-finger (B-reader) of TFIIB for eukaryotic Pol II (63,64) and the linker of Brf1 for eukaryotic Pol III (Supplementary Figure S6) (65,66). Apparently, distinct multiple-subunit DNA-dependent RNAP have evolved non-homologous, but functionally equivalent structure modules for efficient transcription initiation, implicating a unified mechanism for transcription initiation for multiple-subunit DNA-dependent RNAP.

DATA AVAILABILITY

Atomic coordinates and structure factors of *Ec* σ^E–RPo complex, *Mtb* σ^H–RPo complex, *Mtb* $\sigma^{H/E}$–RPo complex have been deposited into the Protein Data Bank with accession code 6JBQ, 6JCX and 6JCY, respectively (https://www.wwpdb.org/).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Prof. Richard Ebright for generous gifts of pACYC-duet-*Mtb-rpoA-rpoD*, pETDuet-*Mtb-rpoB-rpoC*, Prof. Bryce Nickels for generous gift of pET28c-*Ec*σ^{70}, Prof. Xiaoming Zhang for generous gift of *M. tuberculosis* genomic DNA and Tolo Biotechnology for generous

gift of pTolo-EX vectors. We thank the staff at beamline BL18U1/BL19U1 of National Center for Protein Science Shanghai (NCPSS), and at beamline BL17U1 of Shanghai Synchrotron Radiation Facility for assistance during data collection. We thank Shenghai Chang at center of cryo Electron Microscopy for help with cryo-EM sample preparation and data collection. We thank the state key laboratory of bio-organic and natural products chemistry at Shanghai institute of organic chemistry at CAS for sharing the stopped-flow fluorescence spectrometer.

FUNDING

Strategic Priority Research Program of the Chinese Academy of Sciences [XDB29020000]; National Natural Science Foundation of China [31670067, 31822001]; Leading Science Key Research Program of the Chinese Academy of Sciences [QYZDB-SSW-SMC005]. Funding for open access charge: Chinese Academy of Sciences (QYZDB-SSW-SMC005).
Conflict of interest statement. None declared.

REFERENCES

1. Feklistov,A., Sharon,B.D., Darst,S.A. and Gross,C.A. (2014) Bacterial sigma factors: a historical, structural, and genomic perspective. *Annu. Rev. Microbiol.*, **68**, 357–376.
2. Saecker,R.M., Record,M.T. Jr and Dehaseth,P.L. (2011) Mechanism of bacterial transcription initiation: RNA polymerase—promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. *J. Mol. Biol.*, **412**, 754–771.
3. Paget,M.S. (2015) Bacterial sigma factors and anti-sigma factors: structure, function and distribution. *Biomolecules*, **5**, 1245–1265.
4. Ruff,E.F., Record,M.T. Jr and Artsimovitch,I. (2015) Initial events in bacterial transcription initiation. *Biomolecules*, **5**, 1035–1062.
5. Gruber,T.M. and Gross,C.A. (2003) Multiple sigma subunits and the partitioning of bacterial transcription space. *Annu. Rev. Microbiol.*, **57**, 441–466.
6. Mascher,T. (2013) Signaling diversity and evolution of extracytoplasmic function (ECF) sigma factors. *Curr. Opin. Microbiol.*, **16**, 148–155.
7. Haugen,S.P., Berkmen,M.B., Ross,W., Gaal,T., Ward,C. and Gourse,R.L. (2006) rRNA promoter regulation by nonoptimal binding of sigma region 1.2: an additional recognition element for RNA polymerase. *Cell*, **125**, 1069–1082.
8. Feklistov,A., Barinova,N., Sevostyanova,A., Heyduk,E., Bass,I., Vvedenskaya,I., Kuznedelov,K., Merkiene,E., Stavrovskaya,E., Klimasauskas,S. *et al.* (2006) A basal promoter element recognized by free RNA polymerase sigma subunit determines promoter recognition by RNA polymerase holoenzyme. *Mol. Cell*, **23**, 97–107.
9. Narayanan,A., Vago,F.S., Li,K., Qayyum,M.Z., Yernool,D., Jiang,W. and Murakami,K.S. (2018) 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.
10. Zhang,Y., Feng,Y., Chatterjee,S., Tuske,S., Ho,M.X., Arnold,E. and Ebright,R.H. (2012) Structural basis of transcription initiation. *Science*, **338**, 1076–1080.
11. Feklistov,A. and Darst,S.A. (2011) Structural basis for promoter-10 element recognition by the bacterial RNA polymerase sigma subunit. *Cell*, **147**, 1257–1269.
12. Lane,W.J. and Darst,S.A. (2006) The structural basis for promoter -35 element recognition by the group IV sigma factors. *PLoS Biol.*, **4**, e269.
13. Bae,B., Feklistov,A., Lass-Napiorkowska,A., Landick,R. and Darst,S.A. (2015) Structure of a bacterial RNA polymerase holoenzyme open promoter complex. *Elife*, **4**, doi:10.7554/eLife.08504.
14. Feng,Y., Zhang,Y. and Ebright,R.H. (2016) Structural basis of transcription activation. *Science*, **352**, 1330–1333.
15. Juang,Y.L. and Helmann,J.D. (1994) A promoter melting region in the primary sigma factor of Bacillus subtilis. Identification of functionally important aromatic amino acids. *J. Mol. Biol.*, **235**, 1470–1488.
16. Basu,R.S., Warner,B.A., Molodtsov,V., Pupov,D., Esyunina,D., Fernandez-Tornero,C., Kulbachinskiy,A. and Murakami,K.S. (2014) Structural basis of transcription initiation by bacterial RNA polymerase holoenzyme. *J. Biol. Chem.*, **289**, 24549–24559.
17. Kulbachinskiy,A. and Mustaev,A. (2006) Region 3.2 of the sigma subunit contributes to the binding of the 3′-initiating nucleotide in the RNA polymerase active center and facilitates promoter clearance during initiation. *J. Biol. Chem.*, **281**, 18273–18276.
18. Lerner,E., Chung,S., Allen,B.L., Wang,S., Lee,J., Lu,S.W., Grimaud,L.W., Ingargiola,A., Michalet,X., Alhadid,Y. *et al.* (2016) Backtracked and paused transcription initiation intermediate of Escherichia coli RNA polymerase. *PNAS*, **113**, E6562–E6571.
19. Duchi,D., Bauer,D.L., Fernandez,L., Evans,G., Robb,N., Hwang,L.C., Gryte,K., Tomescu,A., Zawadzki,P., Morichaud,Z. *et al.* (2016) RNA polymerase pausing during initial transcription. *Mol. Cell*, **63**, 939–950.
20. Dulin,D., Bauer,D.L.V., Malinen,A.M., Bakermans,J.J.W., Kaller,M., Morichaud,Z., Petushkov,I., Depken,M., Brodolin,K., Kulbachinskiy,A. *et al.* (2018) Pausing controls branching between productive and non-productive pathways during initial transcription in bacteria. *Nat. Commun.*, **9**, 1478.
21. Pupov,D., Kuzin,I., Bass,I. and Kulbachinskiy,A. (2014) Distinct functions of the RNA polymerase sigma subunit region 3.2 in RNA priming and promoter escape. *Nucleic Acids Res.*, **42**, 4494–4504.
22. Cashel,M., Hsu,L.M. and Hernandez,V.J. (2003) Changes in conserved region 3 of Escherichia coli sigma 70 reduce abortive transcription and enhance promoter escape. *J. Biol. Chem.*, **278**, 5539–5547.
23. Rodrigue,S., Provvedi,R., Jacques,P.E., Gaudreau,L. and Manganelli,R. (2006) The sigma factors of Mycobacterium tuberculosis. *FEMS Microbiol. Rev.*, **30**, 926–941.
24. Helmann,J.D. (2016) Bacillus subtilis extracytoplasmic function (ECF) sigma factors and defense of the cell envelope. *Curr. Opin. Microbiol.*, **30**, 122–132.
25. Fiebig,A., Herrou,J., Willett,J. and Crosson,S. (2015) General Stress Signaling in the Alphaproteobacteria. *Annu. Rev. Genet.*, **49**, 603–625.
26. Woods,E.C. and McBride,S.M. (2017) Regulation of antimicrobial resistance by extracytoplasmic function (ECF) sigma factors. *Microbes Infect.*, **19**, 238–248.
27. Staron,A., Sofia,H.J., Dietrich,S., Ulrich,L.E., Liesegang,H. and Mascher,T. (2009) The third pillar of bacterial signal transduction: classification of the extracytoplasmic function (ECF) sigma factor protein family. *Mol. Microbiol.*, **74**, 557–581.
28. Rhodius,V.A., Segall-Shapiro,T.H., Sharon,B.D., Ghodasara,A., Orlova,E., Tabakh,H., Burkhardt,D.H., Clancy,K., Peterson,T.C., Gross,C.A. *et al.* (2013) Design of orthogonal genetic switches based on a crosstalk map of sigmas, anti-sigmas, and promoters. *Mol. Syst. Biol.*, **9**, 702.
29. Zong,Y., Zhang,H.M., Lyu,C., Ji,X., Hou,J., Guo,X., Ouyang,Q. and Lou,C. (2017) Insulated transcriptional elements enable precise design of genetic circuits. *Nat. Commun.*, **8**, 52.
30. Erickson,J.W. and Gross,C.A. (1989) Identification of the sigma E subunit of Escherichia coli RNA polymerase: a second alternate sigma factor involved in high-temperature gene expression. *Genes Dev.*, **3**, 1462–1471.
31. Chassaing,B. and Darfeuille-Michaud,A. (2013) The sigmaE pathway is involved in biofilm formation by Crohn's disease-associated adherent-invasive Escherichia coli. *J. Bacteriol.*, **195**, 76–84.
32. Moreau,P.L. (2014) Protective role of the RpoE (sigmaE) and Cpx envelope stress responses against gentamicin killing of nongrowing Escherichia coli incubated under aerobic, phosphate starvation conditions. *FEMS Microbiol. Lett.*, **357**, 151–156.
33. Ades,S.E. (2008) Regulation by destruction: design of the sigmaE envelope stress response. *Curr. Opin. Microbiol.*, **11**, 535–540.
34. Johansen,J., Rasmussen,A.A., Overgaard,M. and Valentin-Hansen,P. (2006) Conserved small non-coding RNAs that belong to the sigmaE regulon: role in down-regulation of outer membrane proteins. *J. Mol. Biol.*, **364**, 1–8.
35. Guo,M.S., Updegrove,T.B., Gogol,E.B., Shabalina,S.A., Gross,C.A. and Storz,G. (2014) MicL, a new sigmaE-dependent sRNA, combats

envelope stress by repressing synthesis of Lpp, the major outer membrane lipoprotein. *Genes Dev.*, **28**, 1620–1634.
36. Rhodius,V.A., Suh,W.C., Nonaka,G., West,J. and Gross,C.A. (2006) Conserved and variable functions of the sigmaE stress response in related genomes. *PLoS Biol.*, **4**, e2.
37. Rhodius,V.A. and Mutalik,V.K. (2010) Predicting strength and function for promoters of the Escherichia coli alternative sigma factor, sigmaE. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 2854–2859.
38. Dartigalongue,C., Missiakas,D. and Raina,S. (2001) Characterization of the Escherichia coli sigma E regulon. *J. Biol. Chem.*, **276**, 20866–20875.
39. Campagne,S., Marsh,M.E., Capitani,G., Vorholt,J.A. and Allain,F.H. (2014) Structural basis for -10 promoter element melting by environmentally induced sigma factors. *Nat. Struct. Mol. Biol.*, **21**, 269–276.
40. Campbell,E.A., Tupy,J.L., Gruber,T.M., Wang,S., Sharp,M.M., Gross,C.A. and Darst,S.A. (2003) Crystal structure of Escherichia coli sigmaE with the cytoplasmic domain of its anti-sigma RseA. *Mol. Cell*, **11**, 1067–1078.
41. Li,L., Fang,C., Zhuang,N., Wang,T. and Zhang,Y. (2019) Structural basis for transcription initiation by bacterial ECF sigma factors. *Nat. Commun.*, **10**, 1153.
42. Lin,W., Mandal,S., Degen,D., Cho,M.S., Feng,Y., Das,K. and Ebright,R.H. (2019) Structural basis of ECF-sigma-factor-dependent transcription initiation. *Nat. Commun.*, **10**, 710.
43. Vassylyev,D.G., Sekine,S., Laptenko,O., Lee,J., Vassylyeva,M.N., Borukhov,S. and Yokoyama,S. (2002) Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 A resolution. *Nature*, **417**, 712–719.
44. Murakami,K.S., Masuda,S. and Darst,S.A. (2002) Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 A resolution. *Science*, **296**, 1280–1284.
45. Chakraborty,A., Wang,D., Ebright,Y.W., Korlann,Y., Kortkhonjia,E., Kim,T., Chowdhury,S., Wigneshweraraj,S., Irschik,H., Jansen,R. *et al.* (2012) Opening and closing of the bacterial RNA polymerase clamp. *Science*, **337**, 591–595.
46. Feklistov,A., Bae,B., Hauver,J., Lass-Napiorkowska,A., Kalesse,M., Glaus,F., Altmann,K.H., Heyduk,T., Landick,R. and Darst,S.A. (2017) RNA polymerase motions during promoter melting. *Science*, **356**, 863–866.
47. Lin,W., Mandal,S., Degen,D., Liu,Y., Ebright,Y.W., Li,S., Feng,Y., Zhang,Y., Jiang,Y., Liu,S. *et al.* (2017) Structural basis of mycobacterium tuberculosis transcription and transcription inhibition. *Mol. Cell*, **66**, 169–179.
48. Zuo,Y. and Steitz,T.A. (2015) Crystal structures of the E. coli transcription initiation complexes with a complete bubble. *Mol. Cell*, **58**, 534–540.
49. Murakami,K.S., Masuda,S., Campbell,E.A., Muzzin,O. and Darst,S.A. (2002) Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science*, **296**, 1285–1290.
50. Sullivan,J.J., Bjornson,K.P., Sowers,L.C. and deHaseth,P.L. (1997) Spectroscopic determination of open complex formation at promoters for Escherichia coli RNA polymerase. *Biochemistry*, **36**, 8005–8012.
51. Fedoriw,A.M., Liu,H., Anderson,V.E. and deHaseth,P.L. (1998) Equilibrium and kinetic parameters of the sequence-specific interaction of Escherichia coli RNA polymerase with nontemplate strand oligodeoxyribonucleotides. *Biochemistry*, **37**, 11971–11979.
52. Schroeder,L.A., Gries,T.J., Saecker,R.M., Record,M.T. Jr, Harris,M.E. and DeHaseth,P.L. (2009) Evidence for a tyrosine-adenine stacking interaction and for a short-lived open intermediate subsequent to initial binding of Escherichia coli RNA polymerase to promoter DNA. *J. Mol. Biol.*, **385**, 339–349.
53. Wang,S., Li,W., Liu,S. and Xu,J. (2016) RaptorX-Property: a web server for protein structure property prediction. *Nucleic Acids Res.*, **44**, W430–W435.
54. Wang,S., Peng,J., Ma,J. and Xu,J. (2016) Protein secondary structure prediction using deep convolutional neural fields. *Sci. Rep.*, **6**, 18962.
55. Maeda,H., Fujita,N. and Ishihama,A. (2000) Competition among seven Escherichia coli sigma subunits: relative binding affinities to the core RNA polymerase. *Nucleic Acids Res.*, **28**, 3497–3503.
56. Bae,B., Chen,J., Davis,E., Leon,K., Darst,S.A. and Campbell,E.A. (2015) CarD uses a minor groove wedge mechanism to stabilize the RNA polymerase open promoter complex. *Elife*, **4**, doi:10.7554/eLife.08505.
57. Petushkov,I., Esyunina,D., Mekler,V., Severinov,K., Pupov,D. and Kulbachinskiy,A. (2017) Interplay between sigma region 3.2 and secondary channel factors during promoter escape by bacterial RNA polymerase. *Biochem. J.*, **474**, 4053–4064.
58. Glyde,R., Ye,F., Darbari,V.C., Zhang,N., Buck,M. and Zhang,X. (2017) Structures of RNA polymerase closed and intermediate complexes reveal mechanisms of DNA opening and transcription initiation. *Mol. Cell*, **67**, 106–116.
59. Campbell,E.A., Kamath,S., Rajashankar,K.R., Wu,M. and Darst,S.A. (2017) Crystal structure of Aquifex aeolicus sigma(N) bound to promoter DNA and the structure of sigma(N)-holoenzyme. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, E1805–E1814.
60. Dexl,S., Reichelt,R., Kraatz,K., Schulz,S., Grohmann,D., Bartlett,M. and Thomm,M. (2018) Displacement of the transcription factor B reader domain during transcription initiation. *Nucleic Acids Res.*, **46**, 10066–10081.
61. Engel,C., Gubbey,T., Neyer,S., Sainsbury,S., Oberthuer,C., Baejen,C., Bernecky,C. and Cramer,P. (2017) Structural basis of RNA polymerase I transcription initiation. *Cell*, **169**, 120–131.
62. Han,Y., Yan,C., Nguyen,T.H.D., Jackobel,A.J., Ivanov,I., Knutson,B.A. and He,Y. (2017) Structural mechanism of ATP-independent transcription initiation by RNA polymerase I. *Elife*, **6**, e27414.
63. Sainsbury,S., Niesser,J. and Cramer,P. (2013) Structure and function of the initially transcribing RNA polymerase II-TFIIB complex. *Nature*, **493**, 437–440.
64. He,Y., Yan,C., Fang,J., Inouye,C., Tjian,R., Ivanov,I. and Nogales,E. (2016) Near-atomic resolution visualization of human transcription promoter opening. *Nature*, **533**, 359–365.
65. Abascal-Palacios,G., Ramsay,E.P., Beuron,F., Morris,E. and Vannini,A. (2018) Structural basis of RNA polymerase III transcription initiation. *Nature*, **553**, 301–306.
66. Vorlander,M.K., Khatter,H., Wetzel,R., Hagen,W.J.H. and Muller,C.W. (2018) Molecular mechanism of promoter opening by RNA polymerase III. *Nature*, **553**, 295–300.