


ARTICLE

<https://doi.org/10.1038/s41467-019-09096-y>

OPEN

Structural basis for transcription initiation by bacterial ECF σ factors

Lingting Li^{1,2}, Chengli Fang^{1,2}, Ningning Zhuang¹, Tiantian Wang^{1,2} & Yu Zhang¹ ¹

Bacterial RNA polymerase employs extra-cytoplasmic function (ECF) σ factors to regulate context-specific gene expression programs. Despite being the most abundant and divergent σ factor class, the structural basis of ECF σ factor-mediated transcription initiation remains unknown. Here, we determine a crystal structure of *Mycobacterium tuberculosis* (*Mtb*) RNAP holoenzyme comprising an RNAP core enzyme and the ECF σ factor σ^H (σ^H -RNAP) at 2.7 Å, and solve another crystal structure of a transcription initiation complex of *Mtb* σ^H -RNAP (σ^H -RPo) comprising promoter DNA and an RNA primer at 2.8 Å. The two structures together reveal the interactions between σ^H and RNAP that are essential for σ^H -RNAP holoenzyme assembly as well as the interactions between σ^H -RNAP and promoter DNA responsible for stringent promoter recognition and for promoter unwinding. Our study establishes that ECF σ factors and primary σ factors employ distinct mechanisms for promoter recognition and for promoter unwinding.

¹Key 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. ²University of Chinese Academy of Sciences, Beijing 100049, China. These authors contributed equally: Lingting Li, Chengli Fang. Correspondence and requests for materials should be addressed to Y.Z. (email: yzhang@sippe.ac.cn)

Transcription initiation is the first and the most tightly regulated step of bacterial gene expression^{1–3}. σ factors are required for transcription initiation⁴. After forming a complex with the RNA polymerase (RNAP) core enzyme, σ factors guide RNAP to promoter DNA, open double-stranded DNA (dsDNA) to form a transcription bubble, facilitate synthesis of initial short RNA transcripts, and later assist in promoter escape^{4–6}.

Bacterial σ factors are classified into two types— σ^{70} - and σ^{54} -type factors based on their distinct structures and mechanisms. The σ^{70} -type factors can be further classified into four groups according to their conserved domains⁴. Group-1 σ factors (or primary σ factors) contain domains $\sigma_{1.1}$, $\sigma_{1.2}$, σ_{NCR} , σ_2 , $\sigma_{3.1}$, $\sigma_{3.2}$, and σ_4 ; group-2 σ factors contain all domains except $\sigma_{1.1}$; group-3 σ factors contain σ_2 , $\sigma_{3.1}$, $\sigma_{3.2}$, and σ_4 ; while group-4 or extracytoplasmic function (ECF) σ factors only contain σ_2 and σ_4 ⁷. The genomes of a majority of bacteria harbor one primary σ factor for expression of most genes (i.e., group-1 σ factor; σ^{70} in *Escherichia coli* and σ^A in Gram-positive bacteria; referred as σ^A hereafter), and multiple alternative σ factors for expression of genes with cellular- or environmental-context-dependent functions^{8,9}. The ECF σ factors are the largest family of alternative σ factors. On average, bacterial genomes encode six ECF σ factors; the given number for a particular bacterium will vary according to its genome size and environmental complexity^{9,10}. ECF σ factors enable bacteria to rapidly respond to a variety of stresses^{9,11,12} and are known to be essential for the pathogenicity of several disease-causing bacteria^{13,14}. *Mycobacterium tuberculosis* has 10 ECF σ factors (σ^C , σ^D , σ^E , σ^G , σ^H , σ^I , σ^J , σ^K , σ^L , and σ^M); deletion of ECF σ factors from *M. tuberculosis* results in attenuated disease progression (e.g., *sigC* and *sigD*) or in alleviated virulence (e.g., *sigE* and *sigH*)^{15,16}.

The σ^A is capable of recognizing at least five conserved functional elements in the DNA sequences of gene promoters, including the “–35 element” (TTGACA)¹⁷, the “Z element”¹⁸, the “extended –10 element” (TG)¹⁹, the “–10 element” (TATAAT)¹⁷, and the “discriminator element” (GGG)²⁰. Distinct domains of σ^A are responsible for interacting with these DNA elements: the domain σ_4 forms sequence-specific interactions with exposed bases in the major groove of the –35 dsDNA²¹; the $\sigma_{2.5}$ and $\sigma_{3.1}$ domains reach into the major groove of the extended –10 element and make base-specific contacts^{22,23}; and the σ_2 and $\sigma_{1.2}$ domains recognize and then unwind the –10 element dsDNA^{3,24}.

During the process of promoter unwinding, a tryptophan dyad of σ_2 (W256/W257 in *Thermus aquaticus* σ^A or W433/W434 in *E. coli* σ^{70}) forms a chair-like structure that functions as a wedge to separate the dsDNA at the (–12)/(–11) junction^{22,23}. The group-2 σ factors use the same set of residues to unwind promoter DNA; but the melting residues of group-3 σ factors are not conserved²⁵. Subsequently, the base moieties of the unwound nucleotides at position –11 and –7 of the nontemplate strand— $A_{(-11)}(\text{nt})$ and $T_{(-7)}(\text{nt})$ —are flipped out and inserted into preformed pockets by σ_2 and $\sigma_{1.2}$ ^{3,24}. Domain $\sigma_{1.2}$ also recognizes the discriminator element by flipping out the guanine base of $G_{(-6)}(\text{nt})$ and inserting it into a pocket²⁴. Although $\sigma_{3.2}$ does not read the promoter sequence directly, it is essential for transcription initiation. Domain $\sigma_{3.2}$ reaches into the RNAP active site cleft and “pre-organizes” template single-stranded DNA (ssDNA)²⁴. Domain $\sigma_{3.2}$ also blocks the path of the extending RNA chain (>5 nt)^{26,27} thereby contributing to both initial transcription pausing²⁸ and promoter escape^{29,30}.

Each category of known ECF σ factors recognizes promoters bearing a unique sequence signature at the –35 and the –10 elements^{10,31}. In contrast to the high tolerance to sequence variation at the –35 and the –10 promoter elements exhibited by

the primary σ factor, the ECF σ factors have stringent requirements for sequence identity in the –35 and the –10 elements and for spacer length between these two elements through an unknown mechanism^{8,32}. Although both the primary and ECF σ factors recognize the –35 element via σ_4 and recognize the –10 element via σ_2 , the protein sequences of these two domains are not well conserved, and the consensus sequences of the two corresponding DNA elements vary. Crystal structures of individual σ_2 or σ_4 domains of ECF σ factors complexed with cognate DNA have suggested that these ECF domains bind the –35 and the –10 elements differently than does the primary σ factor, implicating a unique means of promoter recognition by ECF σ factors^{33,34}. Another striking difference was revealed by a sequence analysis showing the surprising fact that ECF σ factors do not contain σ_3 domains ($\sigma_{3.1}$ and $\sigma_{3.2}$), but instead contains a linker—highly variable in both length and sequence—to connect the σ_2 and σ_4 domains⁴. This fact immediately raises the question of how these σ factors perform multiple steps of transcription initiation that the σ_3 domain performs in the primary σ factor.

A recent crystal structure of *E. coli* σ^{E_2} –10 ssDNA binary complex suggests that bacterial ECF σ factors probably recognize and unwind promoters through a unique mechanism. Specifically, *E. coli* σ^E employs a flexible “specificity loop” to recognize a flipped master nucleotide of the –10 element and probably unwinds at a distinct position compared with that of σ^{70} by using non-conserved melting residues³⁴. In contrast to the large collection of structural information of primary σ factor, no structure of bacterial RNAP complex with ECF σ factor is available. Therefore, it is largely unknown how ECF σ factors form a holoenzyme with RNAP and how ECF σ factors work alongside RNAP to recognize and to unwind promoter DNA. Here we report the crystal structure of an ECF σ factor–RNAP holoenzyme comprising *M. tuberculosis* RNAP and σ^H at 2.70 Å resolution. We also report the crystal structure of an ECF σ factor–RNAP transcription initiation complex comprising *M. tuberculosis* σ^H –RNAP holoenzyme, a full transcription bubble of promoter DNA, and an RNA primer at 2.80 Å resolution. The crystal structures present detailed interactions among RNAP, ECF σ factors, and promoter DNA. The structures together with data from biochemical assays collectively establish the structural basis of RNAP holoenzyme assembly, promoter recognition, and promoter unwinding by the ECF σ factors.

Results

The crystal structure of *M. tuberculosis* σ^H –RNAP holoenzyme.

The crystals of *M. tuberculosis* σ^H –RNAP holoenzyme were unexpectedly obtained during an initial attempt to crystallize *M. tuberculosis* σ^H –RPO (Supplementary Figs. 1a, c–f). The crystal structure of *M. tuberculosis* σ^H –RNAP holoenzyme at 2.7 Å resolution was determined by molecular replacement using a *Mycobacterium smegmatis* RNAP core enzyme (PDB: 5TW1) [<https://www.rcsb.org/structure/5TW1>] as the searching model³⁵. The Fo–Fc map shows unambiguous density for σ^H residues 22–195 (Table 1; Supplementary Fig. 2a) and the anomalous difference map shows clear density for 4 out of 5 Se atoms, validating the σ^H model (Supplementary Fig. 2a).

σ^H_2 (residues 22–99) and σ^H_4 (residues 140–195) fold into independent helical domains (Fig. 1a, c). Lacking the $\sigma_{1.1}$, $\sigma_{1.2}$, and σ_{NCR} domains of σ^A , the σ^H_2 domain is very compact, containing only four α helices (Fig. 1c). The “specificity loop” (residues 72–79 in σ^H_2 ; Supplementary Fig. 2a) known to be essential for recognition of the –10 element is disordered (no electron density), in contrast to the pre-organized specificity loop in the σ^A –RNAP holoenzyme (Fig. 1b, d and Supplementary Fig. 2b–c). Lacking $\sigma_{1.2}$, the domain that forms extensive

Table 1 The statistics of crystal structures

	<i>M. tuberculosis</i> $\sigma^{\text{H-RNAP}}$	<i>M. tuberculosis</i> $\sigma^{\text{H-RPo}}$
<i>Data collection</i>		
Space group	$P2_1$	$P2_12_1$
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	130.6, 159.8, 131.4	129.8, 164.0, 214.8
α , β , γ (°)	90, 119, 90	90, 90, 90
Resolution (Å)	50.00–2.75 (2.80–2.75)	50.00–2.80 (2.85–2.80)
R_{sym} or R_{merge}	0.078 (0.893)	0.164 (1.430)
$I/\sigma I$	21.4 (1.2)	12.9 (1.2)
Completeness (%)	98.2 (84.3)	99.8 (99.5)
Redundancy	5.4 (3.5)	8.5 (8.4)
CC _{1/2} in highest shell	0.530	0.547
<i>Refinement</i>		
Resolution (Å)	50.00–2.75	50.00–2.80
No. of reflections	119,643	112,969
$R_{\text{work}}/R_{\text{free}}$	0.218/0.258	0.220/0.255
No. of atoms	23,311	25,858
<i>B</i> -factors (Å ²)	90.5	66.1
R.m.s deviations		
Bond lengths (Å)	0.003	0.005
Bond angles (°)	0.565	0.621
Ramachandran plot		
Favored (%)	97.63	97.40
Allowed (%)	2.37	2.60
Disallowed (%)	0	0

Numbers in parenthesis are for the highest resolution

interactions with the specificity loop of σ^{A} , probably accounts for the disordered conformation of the specificity loop in σ^{H} (Fig. 1c, d). As occurs in σ^{A} , the σ^{H_2} domain resides in a cleft between the RNAP- β lobe and the RNAP- β' coiled-coil (β' CC) and makes extensive electrostatic interactions with the latter (Fig. 1e). Notably, the residues contacting β' CC of both σ^{A} and ECF σ factors are conserved (Supplementary Figs. 2d, e and 3), suggesting that β' CC probably serves as an anchor point for the σ_2 domain of the majority of bacterial σ factors.

The σ^{H_4} domain enfolds the flap-tip helix of the RNAP- β subunit (β FTH; Figs. 1e and 2a). The hydrophobic residues contacting the β FTH are conserved between the σ^{A} and the ECF σ factors (Supplementary Figs. 2f–g and 3). Surprisingly, we discovered another anchor point for the σ^{H_4} domain on RNAP—a C-terminal helix of the RNAP- β subunit (β CTH; residues 1145–1157; Figs. 1e and 2a, and Supplementary Fig. 2h and m). The interaction with β CTH was not observed in any of the previously reported bacterial σ^{A} -RNAP structures^{22,24,36,37}. To explore the contribution of such interaction to the transcription activity of σ^{H} -RNAP, we performed in vitro transcription experiments using wild-type or β CTH-deleted *Mtb* σ^{H} -RNAP holoenzyme and *pClpB* promoter variants with –35/–10 spacer lengths ranging from 15 to 19 bp. The wild-type σ^{H} -RNAP was most transcriptionally active with a promoter of 17-bp spacer (Fig. 2b), consistent with a study reporting that most σ^{H} -regulated promoters have a 17-bp spacer³⁸. The β CTH-deletion variant caused impaired transcription activity from promoter with the optimal spacer length (17 bp) but showed little effect on promoter with sub-optimal spacer lengths (16 and 18 bp) (Fig. 2b), suggesting that the interactions between β CTH and σ^{H} are important for the transcription activity of σ^{H} . Intriguingly, deletion of β CTH caused a general increase of σ^{A} -dependent transcription activity from promoter with spacer lengths 15–19 bp (Supplementary Fig. 2i).

The most surprising finding in the structure of σ^{H} -RNAP is the interaction between RNAP and the linker region connecting σ^{H_2}

and σ^{H_4} . The linker is the least conserved region among the ECF σ factors and shares no sequence similarity with the linker of σ^{A} (Supplementary Fig. 3). Our structure shows that the linker region of σ^{H} dives into the active site cleft and emerges out from the RNA exit channel of RNAP (Figs. 1e and 2c, and Supplementary Fig. 4a–b). This interaction creates an entry channel for template ssDNA loading into the active site cleft during RPo formation, but blocks the exit pathway of extending RNA during subsequent transcription initiation events. The path of the $\sigma^{\text{H}_2}/\sigma^{\text{H}_4}$ linker in RNAP is similar to that of $\sigma^{\text{A}_{3.2}}$ (Supplementary Fig. 4c–d), so we designated the $\sigma^{\text{H}_2}/\sigma^{\text{H}_4}$ linker as $\sigma^{\text{H}_{3.2}}$.

To examine the significance of the interaction between $\sigma^{\text{H}_{3.2}}$ and RNAP, we tested the in vitro transcription activity of RNAP holoenzyme comprising σ^{H} variants with the $\sigma^{\text{H}_{3.2}}$ domain either deleted or swapped. Deleting $\sigma^{\text{H}_{3.2}}$ or replacing $\sigma^{\text{H}_{3.2}}$ with a protein sequence known to be disordered completely abolished the transcription activity of σ^{H} (lanes II and III in Fig. 2d), indicating that $\sigma^{\text{H}_{3.2}}$ does not simply serve as a $\sigma^{\text{H}_2}/\sigma^{\text{H}_4}$ linker; rather, the interaction between $\sigma^{\text{H}_{3.2}}$ and the active site cleft of RNAP is essential for transcription. Interestingly, replacing $\sigma^{\text{H}_{3.2}}$ with the σ_2/σ_4 linker from other ECF σ factor partially recovered the transcription activity (lanes IV, V, and VI in Fig. 2d and Supplementary Fig. 4e). Based on these results, we infer that other ECF σ factors probably have a functional $\sigma_{3.2}$ domain that, while divergent in sequence, likely binds RNAP in a way somehow analogous to $\sigma^{\text{H}_{3.2}}$.

The overall structure of *Mtb* σ^{H} -RPo. To understand how σ^{H} -RNAP holoenzyme recognizes promoter DNA and initiates transcription, we sought to determine a crystal structure of σ^{H} -RPo. We assembled the complex by incubating the RNAP core enzyme, σ^{H} , and a synthetic nucleic-acid scaffold (Fig. 3d, and Supplementary Fig. 1b, 1g–j). The synthetic scaffold comprises an upstream DNA duplex (–34 to –10 with respect to transcription start site at +1) with a consensus –35 element (GGAACA), a non-complimentary transcription bubble (–9 to +2) with a consensus –10 element (GTT), a 7-nt RNA primer complementary to template DNA (–6 to +1), and a downstream DNA duplex (+3 to +13). We determined the crystal structure of σ^{H} -RPo at 2.8 Å resolution by molecular replacement using the crystal structure of *Mtb* σ^{H} -RNAP holoenzyme as a search model (Table 1). The Fo–Fc map contoured at 2.5 σ shows clear density for all nucleotides of nontemplate ssDNA, template ssDNA, and RNA primer of the transcription bubble, as well as for all the nucleotides of the upstream and downstream DNA duplexes (Fig. 3e).

In the σ^{H} -RPo structure, σ^{H} makes the same interactions with RNAP as in the structure of σ^{H} -RNAP holoenzyme. The RNAP clamp adopts a closed conformation as in σ^{A} -RPo^{24,39}, consistent with previous single-molecule fluorescence resonance energy transfer results⁴⁰ and supporting the idea that clamp closure is also an obligatory step of RPo formation in ECF σ -mediated transcription initiation (Fig. 3b). The DNA/RNA hybrid resides in the active site cleft in a post-translocation state, and the downstream DNA duplex is accommodated in the main channel. The conformations of the DNA/RNA hybrid and downstream DNA duplex in σ^{H} -RPo and σ^{A} -RPo are similar (Fig. 3c)²⁴.

The σ^{H} -RPo structure revealed multiple interactions responsible for promoter recognition and promoter unwinding by σ^{H} -RNAP that we will describe in detail in each of the subsequent sections of our manuscript. These include: (1) σ^{H_4} inserts into the major groove and reads the sequence of the –35 element (Figs. 3a and 4a, and Supplementary Fig. 6a); (2) the RNAP- β' subunit stabilizes the upstream DNA duplex by contacting the phosphate backbone of nucleotides at positions –23, –20, and –19 (Figs. 3a

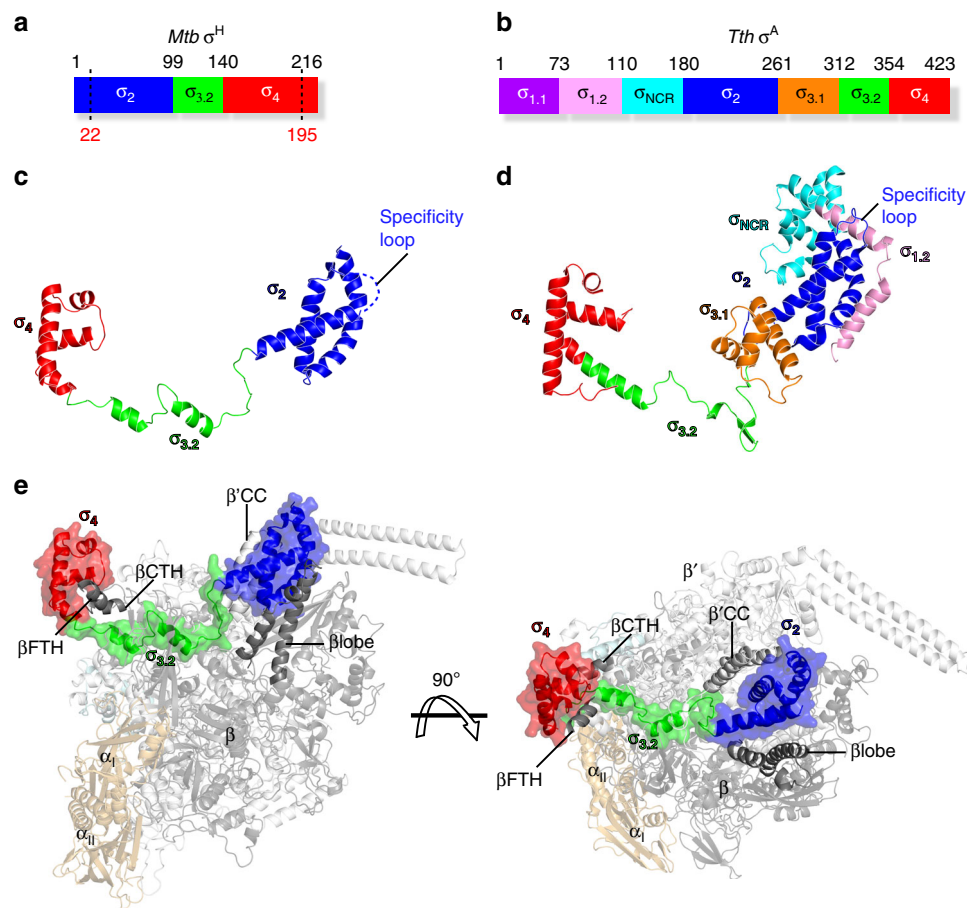


Fig. 1 The crystal structure of *Mtb* σ^H -RNAP holoenzyme. **a** Schematic diagram of *Mtb* σ^H . Ordered regions in the structure are indicated by dashes. **b** Schematic diagram of *T. thermophilus* σ^A . **c** σ^H in the crystal structure of *Mtb* σ^H -RNAP holoenzyme. The disordered specificity loop is shown by blue dashes. **d** σ^A in the crystal structure of *T. thermophilus* σ^A -RNAP (PDB: 1IW7) [<https://www.rcsb.org/structure/1iw7>] was used for comparison due to no available structure of *Mtb* σ^A -RNAP holoenzyme. **e** Front and top views of *Mtb* σ^H -RNAP holoenzyme. β FTH, the flap-tip helix on RNAP- β subunit; β CTH, the C-terminal helix on RNAP- β subunit; β' CC, the coiled-coil on RNAP- β' subunit. $\sigma_{1,1}^A$, purple; $\sigma_{1,2}^A$, pink; σ_{NCR}^A , cyan; σ_2 , blue; $\sigma_{3,1}^A$, orange; $\sigma_{3,2}^A$, green; σ_4 , red. RNAP- α subunits, light orange; RNAP- β subunit, black; RNAP- β' subunit, gray; RNAP- ω subunit, light cyan

and 4b); (3) σ^H_2 unwinds dsDNA using an apparently distinct mechanism (Figs. 3a and 4e); (4) σ^H_2 and the RNAP- β subunit recognize sequences at four positions in the -10 element via interactions with nontemplate ssDNA (Figs. 3a and 5a–c); (5) the RNAP- β subunit recognizes the “CRE element” DNA sequence via interactions with nontemplate ssDNA (Figs. 3a and 5d, and Supplementary Fig. 7a); and (6) $\sigma^H_{3,2}$ guides the template ssDNA into the RNAP active center and forms interactions with the DNA/RNA hybrid in the active site cleft (Figs. 3a and 5e, and Supplementary Fig. 7g).

The interactions of σ^H_4 with the -35 element. σ^H -regulated promoters have a distinct consensus sequence at their -35 elements (5'-GGAAAYA-3'; from -34 to -29 ; Supplementary Fig. 5a)^{38,40,41}. Alternation of DNA sequences at each of the positions from -34 to -29 resulted in substantial loss of transcription activity (Supplementary Fig. 5b). In the structure, σ^H_4 binds to the major groove of dsDNA of the -35 element and makes base-specific polar interactions with nucleotides at three (-34 , -33 , and -31) out of six positions (Figs. 3d and 4a, and Supplementary Fig. 6a). The $G_{-34}(\text{nt})$ makes two H-bonds with R186 through its O6 and N7 atoms; the $G_{-33}(\text{nt})$ makes one H-bond with S182 through its O6 atom; and the $A_{-31}(\text{nt})$ makes one H-bond with M181. Moreover, M181 forms extensive van der Waals interactions with nucleotides at positions -31 and -30 of

the template strand. The interactions are important, as alanine substitutions of R186 or S182 resulted in substantial loss of transcription activity (Fig. 4c). Interestingly, the M181A mutant had increased transcription activity, but this came at the apparent expense of relaxing its sequence stringency for the positions -31 and -30 (Fig. 4d), suggesting that M181 partially accounts for sequence specificity of the two positions.

The rest of the -35 element (-32 , -30 , and -29) makes no base-specific interactions. Previous crystal structure of *E. coli* $\sigma^{E_4}/-35$ dsDNA and a structural model of *Streptomyces coelicolor* $\sigma^{R_4}/-35$ dsDNA reported a local DNA shape readout (straight helix with a narrow minor groove) at this region^{33,42}. We observed a similar DNA conformation (Supplementary Fig. 6b) in our crystal structure, suggesting a general mode of promoter recognition for the ECF σ factors. Such DNA structure might assist the binding of -35 dsDNA to the σ^H_4 surface perhaps by making favorable interactions through its phosphate backbones with polar residues of σ^H_4 including Y166, K167, T179, R183, H185, and R188 (Fig. 3d and Supplementary Fig. 6a). Consistently with this, losing any of these interactions causes a substantial loss of transcription activity (Fig. 4c).

The interactions of σ^H -RNAP with the $-35/-10$ spacer. In the crystal structure of *Mtb* σ^H -RPO, σ^H -RNAP contacts phosphate

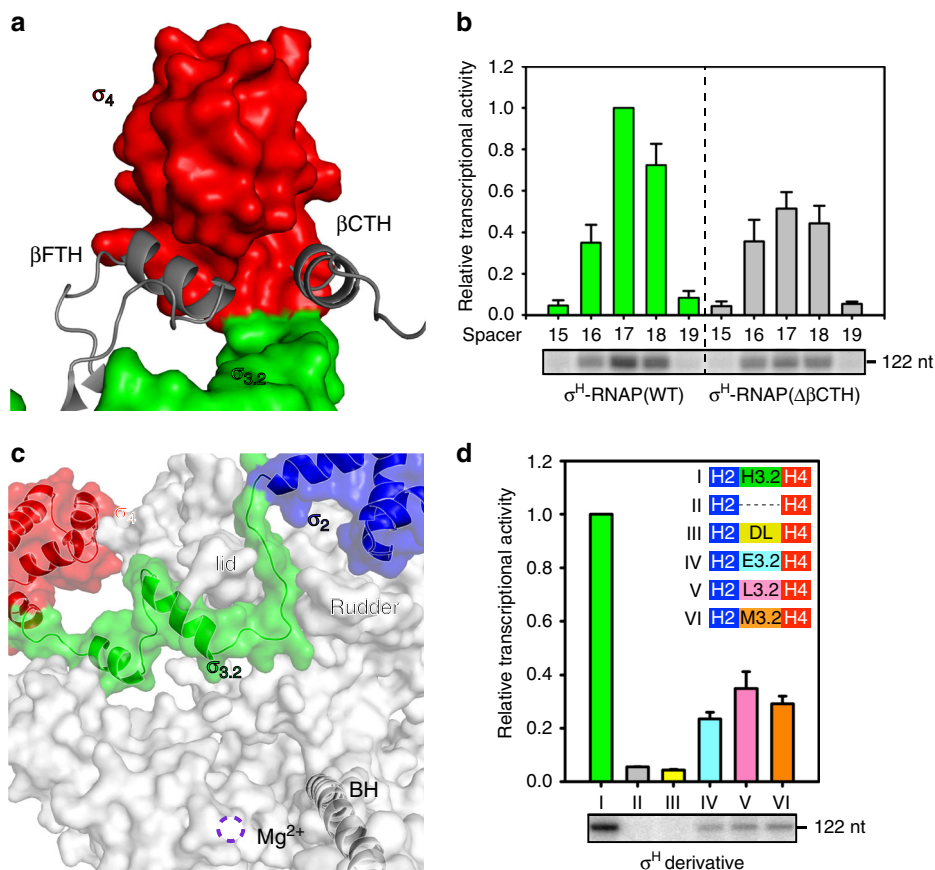


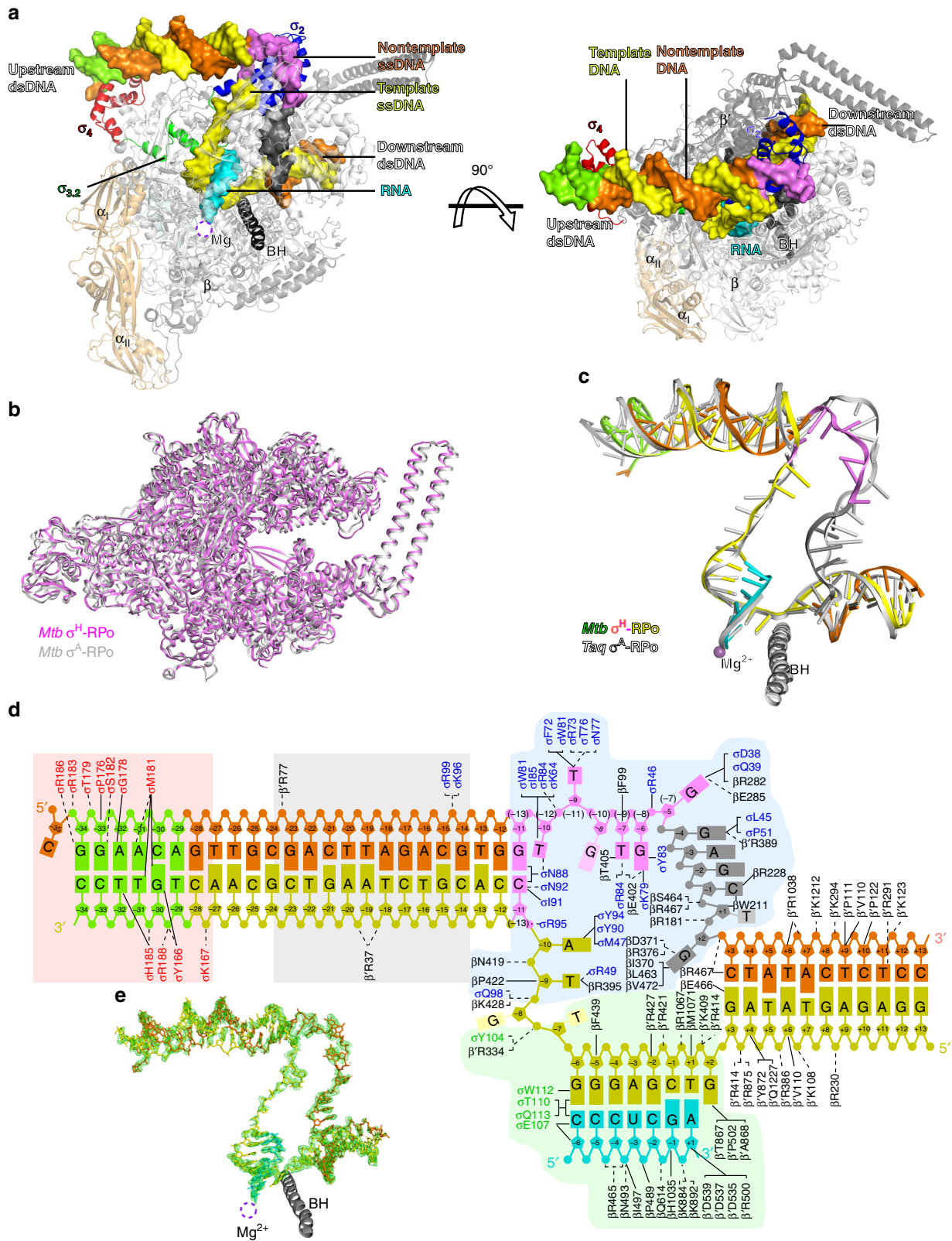
Fig. 2 The interaction between *Mtb* RNAP core enzyme and σ^H . **a** Both β FTH and β CTH interact with σ^H_4 . Colors are as in previous figure. **b** The in vitro transcription activity of σ^H -RNAP(WT) (green bars) and σ^H -RNAP($\Delta\beta$ CTH) (gray bars) from *pClpB* promoter variants with $-35/-10$ spacer length of 15–19 base pairs. “122 nt” indicates length of run-off RNA products. **c** The interaction between σ^H_2/σ^H_4 linker ($\sigma^H_{3.2}$) and RNAP core enzyme. RNAP- β subunit was omitted for clarity. The location of catalytic Mg^{2+} is shown by a dashed purple circle. BH, bridge helix. RNAP- β' subunit, gray. σ^H_2 , $\sigma^H_{3.2}$, and σ^H_4 , blue, green, and red respectively. **d** The in vitro transcription activity from *pClpB* promoter of RNAP holoenzymes comprising σ^H derivatives. H2-H3.2-H4, wild-type σ^H ; H2---H4, two individual domains of σ^H_2 and σ^H_4 ; H2-DL-H4, a chimeric σ^H with $\sigma^H_{3.2}$ replaced by a disordered loop with an equivalent residue number; H2-E3.2-H4, a chimeric σ^H with $\sigma^H_{3.2}$ replaced by *Mtb* $\sigma^E_{3.2}$; H2-L3.2-H4, a chimeric σ^H with $\sigma^H_{3.2}$ replaced by *Mtb* $\sigma^L_{3.2}$; H2-M3.2-H4, a chimeric σ^H with $\sigma^H_{3.2}$ replaced by *Mtb* $\sigma^M_{3.2}$. The experiments were repeated in triplicate, and the data are presented as mean \pm S.E.M. Source data of **b** and **c** are provided as a Source Data file

backbones of the spacer region between the -35 and the -10 elements at three positions (Fig. 3d): (1) R77 of the RNAP- β' zinc-binding domain contacts the nucleotide at position -23 ; (2) R37 of the RNAP- β' zipper domain contacts the nucleotides at positions -20 and -19 ; and (3) K96 and R99 of σ^H_2 contact the nucleotide at position -14 (Fig. 4b). These interactions probably stabilize the conformation of the upstream DNA duplex, likely promoting the engagement of the upstream duplex with σ^H_4 and σ^H_2 for subsequent promoter unwinding. Mutating K96 and R99 causes a mild loss of transcription activity, suggesting the importance of these interactions (Fig. 4f).

The promoter DNA unwinding function of σ^H . The electron density map unambiguously shows that the T:A base pair at position -10 (corresponding to position -12 of promoters for σ^A) is unwound, despite the fact that the -10 nucleotides in the synthetic nucleotide scaffold were designed to be complimentary (Figs. 3d and 4e). This observation strongly suggests that σ^H_2 unwinds promoter DNA at the $-11/-10$ junction (corresponding to $(-13)/(-12)$ of promoters targeted by σ^A ; Fig. 4e and Supplementary Fig. 6c, d). This clearly confirms the hypothesis that the ECF σ factors unwind promoter DNA starting from a distinct position as compared to σ^A ^{32,34}. In the structure, N88 blocks the

pathway of upstream dsDNA and serves as a wedge to disrupt the stacking of base pairs at the positions -11 and -10 . The base pair at position -10 is subsequently forced open by σ N88 via a competitive H-bond between the Watson-Crick atom of the $T_{-10}(\text{nt})$ and N88 (Fig. 4e). Two unwound nucleotides on the nontemplate strand DNA ($T_{-10}(\text{nt})$ and $T_{-9}(\text{nt})$) are stabilized by two adjacent pockets of σ^H_2 ; these pockets are where the sequence identities are “read” (Supplementary Fig. 6F). Moreover, two unwound nucleotides on the template strand DNA ($A_{-10}(\text{t})$ and $T_{-9}(\text{t})$) are also trapped in a cleft created by the RNAP- β lobe and σ^H_2 . Specifically, the base moieties of $A_{-10}(\text{t})$ and $T_{-9}(\text{t})$ form a stack with β R395, σ Y90, and σ Y94, and the phosphate moieties are stabilized by β K428, β N419, and σ Q98. The functional importance of these residues for promoter DNA unwinding was underscored by our finding that their substitution with alanine resulted in defects in transcription (I85G, I85A, and N88A; Fig. 4f).

Structure superimposition between σ^H -RPO and σ^A -RPO shows that the promoter DNA position at which unwinding is initiated differs between σ^A and σ^H by one base pair; σ^H unwinds promoter DNA at a position 1 bp upstream of the position at which σ^A unwinds its promoter DNA (the $(-12)/(-11)$ junction for σ^A ; the $-11/-10$ junction for σ^H corresponding to the $(-13)/(-12)$ junction for σ^A ; Supplementary Fig. 6c–e). A tryptophan



dyad (W433/W434 in *E. coli* or W256/W257 in *T. aquaticus*) is essential for promoter unwinding at the (-12)/(-11) junction by σ^A 22,23,43,44, but the residues at the corresponding positions of σ^H (R84/I85) are not conserved (Supplementary Fig. 3). Mutating R84 and I85 in σ^H to tryptophan (I85W, R84W, or I85W/R84W) resulted in substantial loss of transcription activity, confirming that σ^H opens promoter through a different

mechanism than σ^A and supporting the mechanism proposed for *E. coli* σ^E (Fig. 4f)³⁴.

The interactions of σ^H -RNAP with the -10 element. σ^H -regulated promoters contain a “G₋₁₁T₋₁₀T₋₉” consensus sequence at the -10 element (Supplementary Fig. 5a)^{38,40,41}.

Fig. 3 The crystal structure of *Mtb* σ^H -RPO. **a** Front and top views of σ^H -RPO overall structure. The α , β , β' , and ω subunits of RNA polymerase core enzyme are shown as ribbon and colored in light orange, gray, black, and light cyan respectively. The σ^{H_2} , $\sigma^{H_{3,2}}$, and σ^{H_4} are shown as ribbon and colored as in above figures. The nontemplate DNA, template strand DNA, and RNA strands are shown in surface and colored in orange, yellow, and cyan, respectively, except the -35 element (green), the -10 element (purple), and the CRE (black). The location of the catalytic Mg^{2+} is indicated by a dashed circle. **b** Both *Mtb* σ^H -RPO (violet) and *Mtb* σ^A -RPO (light gray; PDB: 5UHA) [<https://www.rcsb.org/structure/5Uha>] show closed clamp conformation. **c** The comparison of upstream double-stranded DNA (dsDNA), transcription bubble, and downstream dsDNA in *Mtb* σ^H -RPO (colored as in **a**) and *Taq* σ^A -RPO (light gray; PDB: 4XLN) [<https://www.rcsb.org/structure/4XLN>]. **d** Summary of protein-nucleic acid interactions. Solid line, van der Waals interactions; dashed line, polar interactions. Colors are as in above. Red box, interactions with the -35 element (details in Fig. 4a); gray box, interactions with the $-35/-10$ spacer (details in Fig. 4b); blue box, interactions of the single-stranded DNA (ssDNA) in transcription bubble (details in Figs. 4e and 5a-d); green box, interactions with the DNA/RNA hybrid (details in Fig. 5e). The numbers in parenthesis are corresponding positions in σ^A -regulated promoters. **e** The simulated-annealing omit Fo-Fc electron density map (nucleic acids removed; green; contoured at 2.5σ) and model for nucleic acids

Alteration of the DNA sequence at any of these positions resulted in complete loss of promoter activity (Supplementary Fig. 5b), helping explain the reported finding from previous bioinformatic studies that the -10 element is the most conserved region among ECF σ factors^{10,31}. Our crystal structure shows that the base moiety of $C_{-11}(t)$ of the -11 G:C pair makes one H-bond with N92 and extensive Van der Waal interactions with I91, alanine substitution of N92 or I91 resulted in modest or substantial decrease of transcription, respectively (Fig. 5a, f), providing a structural explanation for sequence recognition at this position. Our crystal structure further reveals that σ^H recognizes the nucleotides of the next two positions via two protein pockets on the surface of σ^{H_2} (Fig. 5a).

$T_{-10}(nt)$ is accommodated by a shallow protein pocket on σ^{H_2} wherein I85 forms a stack with the base of $T_{-10}(nt)$ at the pocket bottom and W81 supports the sugar moiety of $T_{-10}(nt)$ on one side of the pocket (Fig. 5a). N88 on the other side of the pocket makes a H-bond with the base moiety of $T_{-10}(nt)$, likely contributing to the sequence specificity known to occur for this position. Alanine substitution of I85 or W81 causes severe defects in transcription activity (Figs. 4f and 5f), emphasizing their importance. Sequence alignment of the 10 *Mtb* ECF σ factors shows that the I85 and W81 are highly conserved (Supplementary Fig. 3), suggesting the $-10(nt)$ pocket probably exists on other ECF σ factors. Alanine substitution of N88 causes defects in transcription activity (Fig. 4f), and the sequence alignment shows that N88 is the most frequent residue at this position. However, other polar residues occur at this position (e.g., a histidine for σ^C and σ^D , and an arginine for σ^I , σ^J , and σ^K) (Supplementary Fig. 3), suggesting that this position may help determine sequence specificity for position -10 of the promoter DNA.

$T_{-9}(nt)$ is flipped out and inserted into a protein pocket formed by the “specificity loop” of σ^{H_2} ³⁴. The thymine base of $T_{-9}(nt)$ stacked on top of W81 makes one H-bond with the main-chain atom of σ^H residue R73 and two H-bonds with the side-chain atoms of σ residues T76 and N77 on the specificity loop (Fig. 3d and Supplementary Fig. 6f). F72 and R73 contact the thymine base via van der Waals interactions. Mutating $T_{-9}(nt)$ completely abolished promoter activity (Supplementary Fig. 5b), and alanine substitution of σ^H residues contacting $T_{-9}(nt)$ (F72A, T76A, N77A, and W81A) severely decreased transcription activity from the consensus promoter (Fig. 5f), verifying the requirement for the $T_{-9}(nt)/\sigma^{H_2}$ interaction for transcription. The fact that both the primary and the ECF σ factors use the specificity loop to read the sequence identity (position -9 of σ^H -regulated promoters corresponding to position (-11) of σ^A -regulated promoters; Supplementary Fig. 6f-h) implies the central importance of this position in promoter DNA^{3,24,34}. Outside of these crucial positions, σ^{H_2} forms fewer interactions with nontemplate nucleotides (positions -8 , -7 , and -6). In the structure, the base moiety of nucleotide at position -8 is

disordered (no electron density) and the base moieties of -7 and -6 nucleotides are sandwiched between residue T405 of the RNAP- β subunit and residue Y83 of σ^{H_2} (Figs. 3d and 5b, and Supplementary Fig. 6i).

A surprising finding in the σ^H -RPO crystal structure is that σ^{H_2} flips the guanine base of $G_{-5}(nt)$ (corresponding to the (-7) position of σ^A -regulated promoters) and inserts into a shallow pocket created by σ^{H_2} and the RNAP- β gate loop (Fig. 5c). In this pocket, $G_{-5}(nt)$ forms a stacking interaction with R282, two H-bonds with E285 on the RNAP- β gate loop, and has van der Waals interactions with D38 and Q39 of σ^{H_2} . Mutations of the RNAP- β gate loop (β R282A or β E285A) cause severely reduced transcription activity (Fig. 5f). As the RNAP- β gate loop makes base-specific contacts to $G_{-5}(nt)$, we tested whether this position exhibits sequence preference. Results of in vitro transcription assays showed that promoters with C or G at this position have much higher transcription activity compared with T or A (Supplementary Fig. 5b), suggesting a sequence preference of $C \sim G > T \sim A$ at such position. Our results therefore show that in σ^H -regulated promoters, the consensus sequence of the -10 elements is extended to “ $G_{-11}T_{-10}T_{-9}N_{-8}N_{-7}N_{-6}S_{-5}$ ”. It is worth noting that σ^A also accommodates the $T_{(-7)}(nt)$ nucleotide in a protein pocket^{3,24}; however, the protein pocket is mainly formed by residues from $\sigma^A_{1,2}$ and σ^A_2 ²⁴, and the structural features that determine sequence specificity are located on σ^A (Supplementary Fig. 6i-k).

Interactions of σ^H -RNAP with the CRE. The guanine base of $G_{+2}(nt)$ is inserted into the “G” pocket in σ^H -RPO (Fig. 5d) and makes essentially the same interaction with residues in the “G” pocket as the $G_{(+2)}(nt)$ does in the σ^A -RPO complex (Figs. 3d and 5d, and Supplementary Fig. 7d-f)²⁴. However, in contrast to the σ^A -RPO, in which the nontemplate $T_{(+1)}(nt)$ forms a stacking interaction with β W211, the nontemplate $T_{+1}(nt)$ in σ^H -RPO was pushed out from the base-stacking. Instead, the nucleotide immediately upstream of $+1$ makes the stacking interaction with β W211 in σ^H -RPO (Fig. 5d and Supplementary Fig. 7a-c). The base moieties of the remaining nucleotides (-4 to -1) between the CRE and -10 elements are stacked in a row between L45 on σ^{H_2} and W211 on the RNAP- β subunit (Fig. 5d and Supplementary Fig. 7a).

Interactions of the template ssDNA. In the σ^H -RPO structure, σ^H and the RNAP- β subunit form an “T-ssDNA entry channel” that guides the template ssDNA into the RNAP active center cleft (Supplementary Fig. 4a). Along the channel, σ^H and RNAP form extensive interactions with the template ssDNA (σ M47, σ R49, σ Y90, σ Y94, σ Q98, σ Y04, β R395, β N419, β P422, β K428, and β' R334; Fig. 3d). Compared to σ^A , which forms extensive interactions between the $\sigma^A_{3,2}$ finger with the ssDNA template nucleotides the active center cleft²⁴, σ^H forms fewer interactions (Fig. 3d

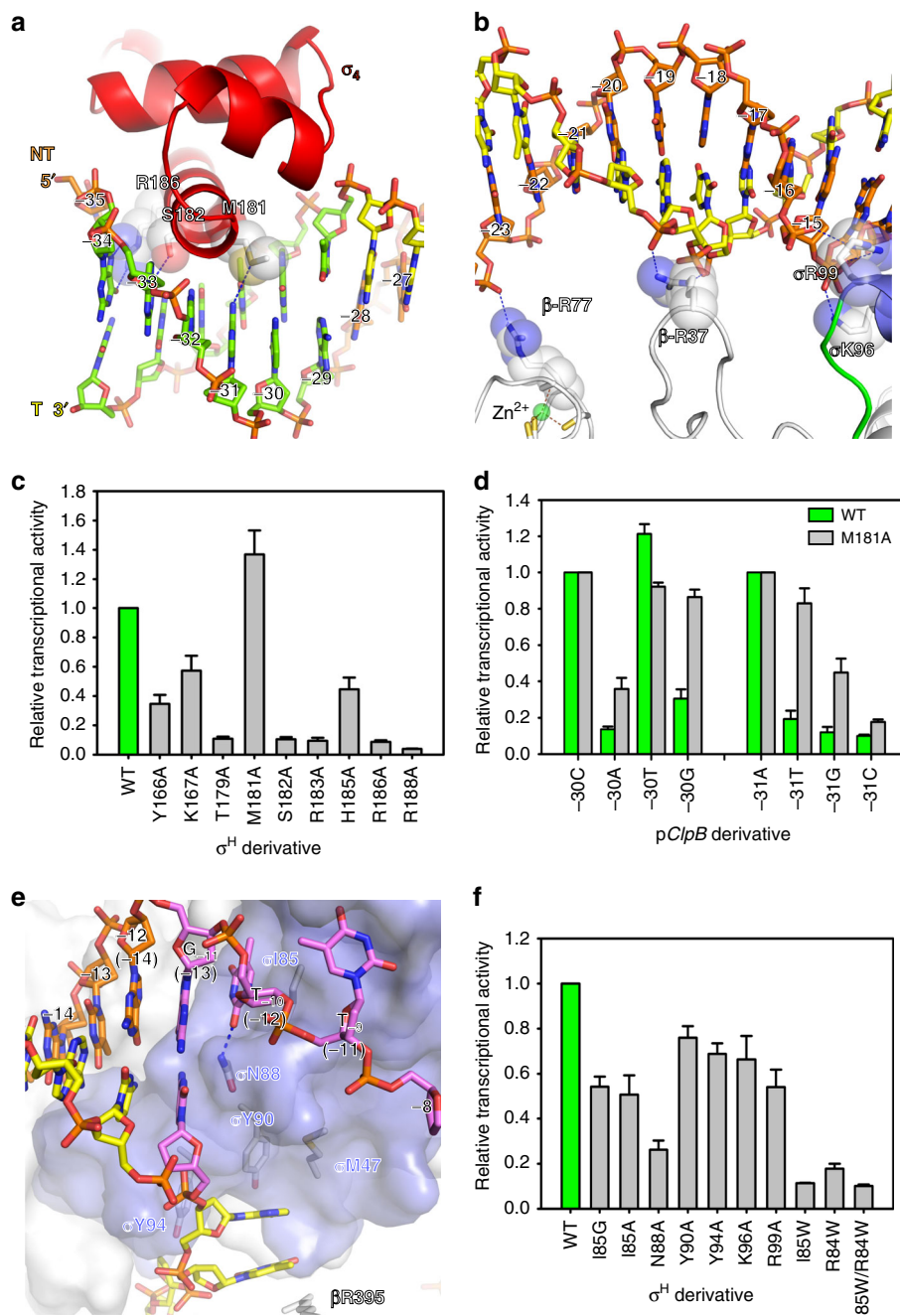


Fig. 4 The interaction between upstream promoter DNA and RNAP in the *Mtb* σ^H -RPo structure. **a** The interaction between σ^H_4 and the -35 element. σ^H_4 , red ribbon. The residues making base-specific interactions are presented as ribbon and half-transparent spheres; the C, O, N, and S atoms of the residues are colored in white, red, blue, and yellow respectively. The C, O, N, and P atoms of the -35 DNA are colored in green, red, blue, and orange, respectively. H-bond, blue dash. **b** The interaction between RNAP and $-35/-10$ spacer region of upstream dsDNA. The residues making polar interactions with DNA are shown and colored as above. The C atoms of the $-35/-10$ spacer DNA are colored in yellow (template DNA) or orange (nontemplate DNA), and the rest of atoms are colored as above. **c** The in vitro transcription activity of σ^H derivatives comprising alanine substitution on σ^H_4 . **d** M181A changes the sequence specificity for positions -30 and -31 of the -35 element. **e** The promoter melting by σ^H . The numbers in parenthesis correspond to positions in σ^A -RPo. The -10 element is colored in purple. The residues involved in promoter melting are shown in stick and colored in white (C atoms), red (O atoms), and blue (N atoms). **f** The in vitro transcription activity of σ^H_2 derivatives with alanine or tryptophan substitutions. The experiments were repeated in triplicate, and the data are presented as mean \pm S.E.M. Source data of **c**, **d**, and **f** are provided as a Source Data file

and Supplementary Fig. 7g–i). The 5' terminus of the 7-nt RNA is positioned very closely to the tip of $\sigma^H_{3,2}$; extending the RNA by even one additional nucleotide would likely result in steric hindrance (Fig. 5e and Supplementary Fig. 7g). Since $\sigma^H_{3,2}$ occupies the RNA exit channel and must be displaced by the nascent RNA chain, such hindrance may be the trigger for the release of σ^H and promoter DNA during promoter escape.

Discussion

The structural basis of transcription initiation by the primary σ factor has been studied extensively, but little is known about how the ECF σ factors—the largest and most diverse group of σ^{70} family factors—initiate transcription. In this study, we present high-resolution crystal structures of *M. tuberculosis* σ^H -RNAP holoenzyme and σ^H -RPo complexes along with comprehensive

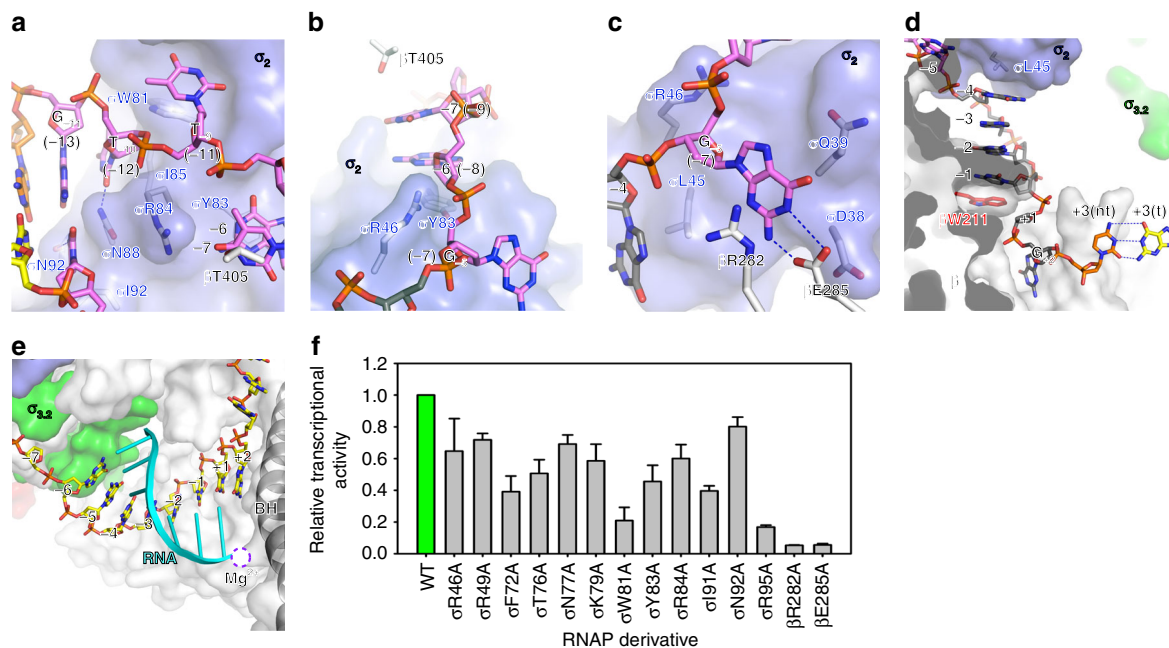


Fig. 5 The interaction between transcription bubble and RNAP in the *Mtb* σ^H -Rpo structure. **a** Recognition of the “G₋₁₁T₋₁₀T₋₉” sequence in the -10 element by σ^H_2 . **b** Stacking of -7(nt) and -6(nt) nucleotides by RNAP- β subunit and σ^H_2 . **c** The recognition of G₋₅(nt) by RNAP- β subunit and σ^H_2 . **d** The interaction between RNA polymerase (RNAP) and CRE element. Colors are as above. **e** The interaction between RNAP and DNA/RNA hybrid. **f** The in vitro transcription activity of RNAP derivatives comprising alanine substitutions of DNA-contacting residues on RNAP- β subunit and σ^H . The experiments were repeated in triplicate, and the data are presented as mean \pm S.E.M. Source data of **f** are provided as a Source Data file

mutational analyses. Our study demonstrates the structural basis for RNAP holoenzyme formation and transcription initiation by the ECF σ factors.

Our structures show that σ^H binds to RNAP in a similar way to σ^A , in which σ_2 and σ_4 stay on the surface of RNAP, and $\sigma_{3,2}$ inserts into the active center. The interactions of σ^H_2 and σ^H_4 with RNAP were thus anticipated and supported a structure model of *E. coli* σ^E -RNAP holoenzyme⁴⁵, as the residues contacting the β FTH and β' CC domains are conserved between the ECF and primary σ factors. However, the interactions of $\sigma^H_{3,2}$ with RNAP are unexpected, showing similarity in neither sequence nor secondary structure between the $\sigma_{3,2}$ regions of σ^{ECF} and σ^A (Supplementary Fig. 3). In vitro transcription experiments show that $\sigma^H_{3,2}$ is essential for the transcription activity of σ^H -RNAP; removing or replacing the linker with an unrelated sequence completely abolished its transcription activity (Fig. 2d). It is worth noting that the B-reader loop of TFIIB reaches into the active site cleft of yeast pol II in a similar way to $\sigma_{3,2}$ ⁴⁶. Considering that this general mode of interaction appears to be conserved between prokaryotic RNAP and eukaryotic pol II, it is reasonable to propose that the σ_2/σ_4 linker of other bacterial ECF σ factors very likely also inserts into the active site cleft of RNAP. Our transcription assays show that chimeric ECF σ factors with swapped $\sigma_{3,2}$ domains retain function in transcription, albeit with reduced activity (Fig. 2d and Supplementary Fig. 4e), supporting this idea. An intriguing question to be answered is how RNAP uses the same channels to accommodate different $\sigma_{3,2}$ domains.

σ^A -Rpo crystal structures show that $\sigma^A_{3,2}$ contacts nucleotides at the template strand of ssDNA and pre-organizes it into an A-form helical conformation in a manner compatible with pairing of initial nucleotide triphosphates (NTPs)^{24,26}. These interactions provide explanations for the effects of $\sigma^A_{3,2}$ on de novo RNA synthesis^{26,47,48}. The structure predicts that the $\sigma^A_{3,2}$ finger has to be displaced by an RNA molecule of >4-nt in length and that the $\sigma^A_{3,2}$ loop in the RNA exit channel has to be cleared during the promoter escape process. The interactions observed in the σ^A -

Rpo structure underscore the key role of the $\sigma^A_{3,2}$ on abortive production, pausing, and promoter escape in transcription initiation^{28,29,49,50}. In our crystal structures of σ^H -Rpo, we show that $\sigma^H_{3,2}$ guides the template ssDNA into the active site cleft and forms interactions with template ssDNA (Figs. 3d and 5e). We propose that $\sigma^H_{3,2}$ probably functions similarly to $\sigma^A_{3,2}$ during transcription initiation: by stabilizing the template ssDNA and facilitating binding of initial NTPs. The crystal structure of σ^H -Rpo also indicates that $\sigma^H_{3,2}$ should collide with RNA molecules >7 nt in length and that the $\sigma^H_{3,2}$ domain must dissociate from the RNAP RNA exit channel during promoter escape (Fig. 5e), raising the possibility that the $\sigma_{3,2}$ of the ECF factors functions like $\sigma^A_{3,2}$ during abortive production, pausing, and promoter escape in transcription initiation.

Our structure of σ^H -Rpo suggests that substantial differences exist between how individual domains σ^H -RNAP and σ^A -RNAP interact with their cognate promoter DNA. Both σ^H_4 and σ^A_4 use the same α -helix to bind the -35 element, but the positions of DNA on the α -helix differ by one α -helical turn, resulting a ~ 4 Å difference in the position of the -35 element on the σ_4 surface (Supplementary Fig. 6B). A previous crystal structure of the *E. coli* σ^E_4 -35 element binary complex is superimposable on our σ^H -Rpo (Supplementary Fig. 6b)³³, suggesting that the distinct mode of interaction that we observed with the -35 element is likely used by other ECF σ factors.

σ^H -RNAP “reads” the sequence of the -10 element differently than does σ^A -RNAP. In our crystal structure of σ^H -Rpo, we discovered that base moieties of three nucleotides—T₋₁₀(nt), T₋₉(nt), and G₋₅(nt) (corresponding to the positions (-12), (-11) and (-7) of σ^A -regulated promoters)—were flipped out and inserted into three respective protein pockets on σ^H (Figs. 3d and 5a-c), in contrast to the two protein pockets known for base moieties of A₍₋₁₁₎(nt) and T₍₋₇₎(nt) on σ^A ^{3,24}. This extra pocket for T₋₁₀(nt) on σ^H was also suggested in a previous structure of a *E. coli* σ^E_2 -10 element binary complex (Supplementary Fig. 6g)³⁴. Sequence alignment of multiple ECF σ factors and σ^A revealed that

residues forming the $T_{-10}(\text{nt})$ pocket are generally conserved between ECF σ factors but are distinct from σ^A , suggesting that other ECF σ factors likely also recognize the nucleotide at this position using similar protein pockets (Supplementary Fig. 3).

σ^H -RNAP uses different protein regions to accommodate the flipped guanine base of $G_{-5}(\text{nt})$ (corresponding to position (-7) of σ^A -regulated promoters) than does σ^A -RNAP for $T_{(-7)}(\text{nt})$ (Supplementary Fig. 6i–k). The guanine base of $G_{-5}(\text{nt})$ is sandwiched between σ^H_2 and the RNAP- β gate loop, while the thymine base of $T_{(-7)}(\text{nt})$ resides in a pocket on $\sigma^A_{1,2}$ ^{3,24}. Our mutation study of the $G_{-5}(\text{nt})$ pocket residues demonstrated that the RNAP- β gate loop functions to recognize this particular nucleotide (Fig. 5f), thus raising the possibility that other σ^{ECF} -RNAP holoenzymes may also bind and read a nucleotide in the nontemplate ssDNA in a manner analogous to σ^H -RNAP.

σ^H -RNAP also engages the -10 element differently than does σ^A -RNAP. We found that the protein pockets for $T_{-9}(\text{nt})$ and $G_{-5}(\text{nt})$ on σ^H -RNAP do not exist in the absence of promoter DNA (Fig. 6a, b). In the crystal structure of σ^H -RNAP, the specificity loop, which recognizes the $T_{-9}(\text{nt})$ is disordered; and the RNAP- β gate loop is too far away from the σ^H_2 to form the $G_{-5}(\text{nt})$ pocket (Fig. 6a, b). Such conformational differences support an “induced-fit” model of interaction between σ^H -RNAP and nontemplate ssDNA, in contrast to the accepted “lock-and-key model” for the interaction between σ^A -RNAP and nontemplate ssDNA (Fig. 6c, d)^{3,24,51}.

σ^H -RNAP recognizes the $G_{+2}(\text{nt})$ of CRE in a same way as does σ^A -RNAP (Supplementary Fig. 7d–f). As the residues that form the “G” pocket are solely from the RNAP core enzyme, it is possible that other ECF σ -RNAP holoenzymes are probably able to read the sequence identity of nucleotide at position $+2$ of the promoter DNA. However, whether the sequence content at this position affects other events (transcription start site selection, slippage synthesis, etc.) during transcription initiation by ECF σ -RNAP as σ^A -RNAP remains to be determined⁵².

Our crystal structures suggest that σ^H employs a distinct mechanism to unwind promoter DNA compared to σ^A (Supplementary Figure 6C–E): (1) σ^H and σ^A use residues with positions that differ by one α -helical turn on the $\sigma_{2,3}$ α -helix (N88 for *Mtb* σ^H vs. W433/W434 for *Ec* σ^A , or W256/W257 for *Taq* σ^A) to unwind promoter DNA; (2) σ^H and σ^A unwind promoter DNA at positions differing by one base pair ($(-13)/(-12)$ junction for σ^H vs. $(-12)/(-11)$ junction for σ^A); and (3) σ^H traps and reads two unwound nucleotides ($T_{(-12)}(\text{nt})$ and $T_{(-11)}(\text{nt})$ immediately after the unwinding points), whereas σ^A only traps and reads one unwound nucleotide ($A_{(-11)}(\text{nt})$). Although it is unclear whether trapping of the flipped nucleotides initiates or facilitates the event of promoter unwinding, such interactions play crucial roles during RPo formation.

Campagne et al. recently identified a similar protein pocket on *E. coli* σ^E_2 for the $T_{(-12)}(\text{nt})$ in a crystal structure of *E. coli* σ^E bound to the -10 element ssDNA, and predicted that *E. coli* σ^E unwinds promoter dsDNA at the $(-13)/(-12)$ junction³⁴. Our crystal structure of σ^H -RPO clearly confirms the unwinding position proposed in the study by Campagne et al.. Sequence alignment of multiple ECF σ factors and σ^A revealed that most of the ECF σ factors do not contain the tryptophan dyad of σ^A at corresponding positions, but instead share a conserved (-12) pocket (Supplementary Figure 3). Therefore, it is possible that the ECF σ factors share the same unwinding mechanism as *Mtb* σ^H and *Ec* σ^E .

Our mutation study of the σ^H -regulated promoter showed that substitution of the consensus sequence at almost every position on the -35 element and -10 element abolished transcription activity (Supplementary Figure 5B). Moreover, extending or shortening the spacer of $-35/-10$ elements substantially reduced

promoter activity (Fig. 2b). These results confirmed previously reported observations that ECF σ factors require a consensus sequence at the $-35/-10$ elements as well as a rigid spacer on promoter DNA to efficiently initiate transcription^{8,32}. Our structures and results from biochemical experiments provide explanations for the promoter stringency of σ^H . We show that the interactions among σ^H and RNAP, the unwinding mechanism, and the induced-fit mode of promoter recognition work in concert to collectively confer the high specificity exhibited by σ^H and probably by other ECF σ factors as well.

We have shown that σ^H employs residues different from σ^A to unwind promoter DNA. The well-conserved tryptophan dyad of σ^A functions very efficiently for promoter unwinding; substitutions of the tryptophan dyad in σ^A resulted in severely reduced transcription activity^{43,44}; and sequence variations at corresponding positions account for inferior DNA unwinding capacity of other alternative σ factors²⁵. Given that ECF σ factors lack the tryptophan dyad at corresponding positions, we infer σ^H (and probably other ECF σ factors) unwinds promoter DNA less efficiently than does σ^A . This putative sub-optimal unwinding efficiency could be compensated by employing a very-high-affinity consensus sequence of promoter DNA to facilitate its loading^{4,8,25}. Our proposed induced-fit mode of nontemplate ssDNA binding by σ^H -RNAP at the position immediate downstream of unwinding—i.e. $T_{-9}(\text{nt})$ —also require the consensus promoter sequence to induce formation of correct conformation of the “specificity loop” (Fig. 6c); RNAP is not able to efficiently propagate promoter unwinding downstream without firmly anchoring the “master” nucleotide— $A_{-11}(\text{nt})$ for *E. coli* σ^{70} corresponding to $T_{-9}(\text{nt})$ for *Mtb* σ^H —as demonstrated in the case of *E. coli* σ^{70} -RNAP^{53–55}.

In conclusion, we demonstrate the structural basis of RNAP holoenzyme formation and transcription initiation by the ECF σ factors, thereby deepening our understanding the basic mechanisms of transcription initiation used by the largest and most diverse group of bacterial initiation factors. Our work will facilitate the rational design of orthogonal transcription units based on ECF σ factors and should help computational chemistry and other efforts to design selective antibacterial agents through the inhibition ECF σ factor-mediated transcription initiation.

Methods

Plasmid construction. The plasmids used in this study are listed in Supplementary 1. For construction of the expression plasmid pTolo-EX5-*Mtb* σ^H , the *M. tuberculosis* σ^H gene amplified from *M. tuberculosis* genomic DNA (see Supplementary Data 1 for primer information) was cloned into the pTolo-EX5 plasmid (Tolo Biotech.) using *NcoI* and *XhoI* restriction sites. The pTolo-EX5-*Mtb* σ^H derivatives bearing single or double mutations were generated through site-directed mutagenesis (Transgen biotech).

The pTolo-EX5-*Mtb* σ^H derivatives encoding chimeric σ^H were generated by replacing the DNA fragment encoding *Mtb* $\sigma^H_{3,2}$ (aa 96–144) with DNA fragments encoding *Ec* σ^A (aa 164–212; disordered acidic loop of the non-conserved region), *Mtb* $\sigma^E_{3,2}$ (aa 150–189), *Mtb* $\sigma^L_{3,2}$ (aa 78–122), or *Mtb* $\sigma^M_{3,2}$ (aa 98–137) in pTolo-EX5-*Mtb* σ^H (Tolo Biotech).

The pACYCDuet-*Mtb-rpoA-rpoZ* plasmid was constructed by replacing *Mtb rpoD* with *Mtb rpoZ* in parent plasmid pACYCDuet-*Mtb-rpoA-sigA* plasmid using *KpnI* and *NdeI* (Supplementary Table 1). The pETduet-*Mtb-rpoB-rpoC* derivatives bearing single mutations were generated through site-directed mutagenesis (Transgen Biotech; Supplementary Table 1 and Supplementary Data 1).

For construction of plasmids for in vitro transcription assays of *Mtb* σ^H , the promoter region (-50 to $+51$) of *ClpB* gene amplified from *M. tuberculosis* genomic DNA was cloned into pEASY-Blunt simple vectors, resulting in pEASY-Blunt-pClpB (Transgen Biotech; Supplementary Table 1 and Supplementary Data 1). The derivatives of pEASY-Blunt-pClpB with varied $-35/-10$ spacer lengths were obtained by site-directed mutagenesis (Supplementary Figure 2J). The promoter region (-50 to $+51$) of *Rv2466c* gene amplified from *M. tuberculosis* genomic DNA was cloned into pEASY-Blunt simple vectors, resulting in pEASY-Blunt-pRv2466c (Supplementary Table 1 and Supplementary Data 1; Supplementary Figure 2L).

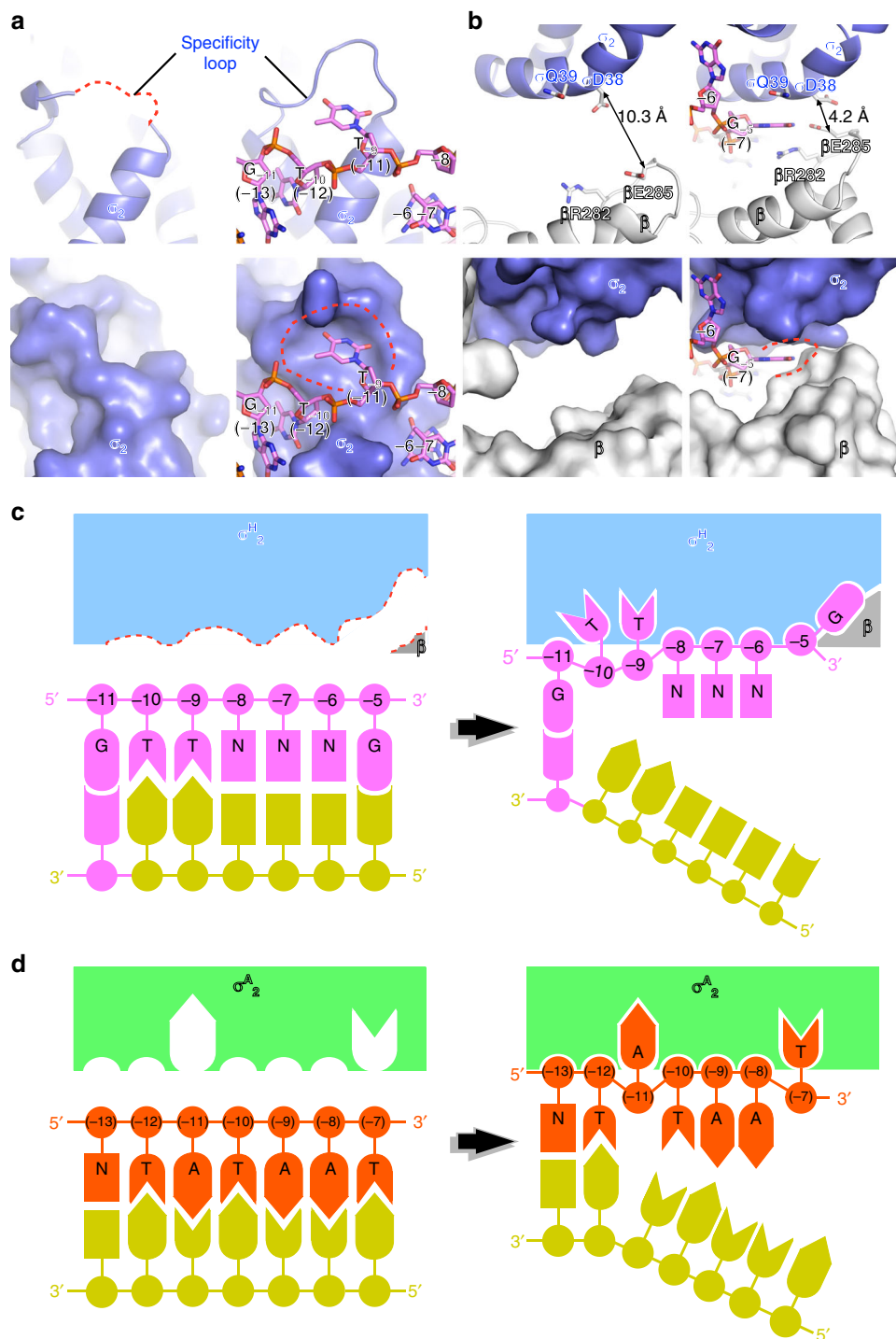


Fig. 6 The induced-fit mechanism of promoter recognition by *Mtb* σ^H -RNAP. **a** The T₋₉(nt) pocket does not exist in σ^H -RNAP holoenzyme (left) but is induced by DNA binding in σ^H -RPo (right) structures. **b** The G₋₅(nt) pocket does not exist in σ^H -RNAP holoenzyme (left) but is induced by DNA binding in σ^H -RPo (right) structures. The pockets are presented in cartoon (top) and surface (bottom). **c** The schematic of induce-fit model of promoter recognition by σ^H -RNAP. **d** The schematic of lock-and-key model of promoter recognition by σ^A -RNAP

The derivatives of pARTaq-N25-100-TR2 for in vitro transcription assays of *Mtb* σ^A with varied -35/-10 spacer lengths were obtained by site-directed mutagenesis (Supplementary Figure 2K).

Protein preparation. For preparation of *M. tuberculosis* σ^H , *E. coli* BL21(DE3) cells (NovoProtein) carrying pTolo-EX5-*Mtb*^H were cultured in Luria-Bertani broth (LB) at 37 °C, and the expression of N-terminal sumo-tagged *Mtb* σ^H was induced at 18 °C for 14 h with 0.5 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) at OD₆₀₀ of 0.8. Cells were harvested by centrifugation (8000 × g, 4 °C), re-

suspended in lysis buffer (20 mM Tris-HCl (pH 8.0), 0.5 M NaCl, 5% (v/v) glycerol, 0.5 mM β -mercaptoethanol, and protease inhibitor cocktail (bimake.cn)) and lysed using an Avestin EmulsiFlex-C3 cell disrupter (Avestin, Inc.). The lysate was centrifuged (16,000 × g; 45 min, 4 °C) and the supernatant was loaded on to a 2 mL column packed with Ni-NTA agarose (SMART, Inc.). The protein was washed by lysis buffer containing 20 mM imidazole and eluted with lysis buffer containing 250 mM imidazole. The eluted fraction was digested by tobacco etch virus protease and dialyzed overnight in dialysis buffer (20 mM Tris-HCl (pH 8.0), 0.2 M NaCl, 1% (v/v) glycerol, and 0.5 mM β -mercaptoethanol). The sample was loaded onto a second Ni-NTA column and the cleaved protein was retrieved from the flow-

through fraction. The sample was diluted to the dialysis buffer with 0.05 M NaCl and further purified through a Heparin column (HiTrap Heparin HP 5 mL column, GE Healthcare Life Sciences) with buffer A (20 mM Tris-HCl (pH 8.0), 0.05 M NaCl, 1% (v/v) glycerol, and 1 mM dithiothreitol (DTT)) and buffer B (20 mM Tris-HCl (pH 8.0), 1 M NaCl, 1% (v/v) glycerol, and 1 mM DTT). Fractions containing *M. tuberculosis* σ^H was concentrated to 5 mg/mL and stored at -80°C . The *M. tuberculosis* σ^H derivatives were prepared by the same procedure.

For preparation of selenomethionines (SeMet)-labeled *M. tuberculosis* σ^H , BL21 (DE3) strains carrying pTolo-EX5-*Mtb* σ^H were cultured in SelenoMet base medium supplemented with nutrient mix (Molecular Dimensions) at 37°C . The amino-acid mixture containing selenomethionine was added into the culture at OD₆₀₀ of 0.4 and the protein expression was induced with 0.5 mM IPTG at OD₆₀₀ of 0.8 for 14 h at 18°C . The SeMet-labeled *Mtb* σ^H was purified as described above.

The *M. tuberculosis* RNAP core enzyme was purified from *E. coli* BL21(DE3) carrying pETDuet-*Mtb-rpoA-rpoZ* and pACYCDuet-*Mtb-rpoB-rpoC* as described⁵⁶. The protein sample was concentrated to 5 mg/mL and stored at -80°C .

Nucleic acid scaffolds. Nucleic acid scaffolds for assembly of σ^H -RNP* for crystallization of σ^H -RNAP holoenzyme was prepared from synthetic oligos (non-template DNA: 5'-GTTGTGCTGGGCGTCACGATGCA-3'; template DNA: 5'-TGCATCCGTGAGTCGGT-3', Sangon Biotech, Supplementary Figure 1A) by an annealing procedure (95°C , 5 min followed by 2°C -step cooling to 25°C) in annealing buffer (5 mM Tris-HCl, pH 8.0, 200 mM NaCl, and 10 mM MgCl₂).

Nucleic acid scaffolds for crystallization of σ^H -RNPo was prepared from synthetic oligos (nontemplate DNA: 5'-CGGAACAGTTGCGACTTAGACGTGGTGTGGGAGCTGCTATACTCTCC-3'; template DNA: 5'-GGAGATATAGTTCGAGGTGTACCAGCTCTAAGTCGCAACTGTTCC-3', Sangon Biotech; and RNA: 5'-CCCUGA-3', Genepharma; Fig. 3c) by an annealing procedure (95°C , 5 min followed by 2°C -step cooling to 25°C) in annealing buffer (5 mM Tris-HCl, pH 8.0, 200 mM NaCl, and 10 mM MgCl₂).

***M. tuberculosis* σ^H -RNPo complex reconstitution.** The *M. tuberculosis* σ^H -RNPo and σ^H -RNP* were reconstituted from *M. tuberculosis* RNAP core enzyme, σ^H (or SeMet- σ^H), and nucleic acid scaffolds. The RNAP core enzyme, σ^H , and nucleic acid scaffolds were mixed at a 1:4:1.2 molar ratio and incubate at 4°C overnight. The mixture was loaded on a HiLoad 16/60 Superdex S200 column (GE Healthcare, Inc.) equilibrated in 20 mM Tris-HCl (pH 8.0), 0.1 M NaCl, 1% (v/v) glycerol, and 1 mM DTT. Fractions containing *Mtb* σ^H -RNPo were collected, concentrated to 7.5 mg/mL, and stored at -80°C .

Structure determination of *M. tuberculosis* σ^H -RNAP holoenzyme. The structure of σ^H -RNAP holoenzyme was obtained during an attempt for obtaining the σ^H -RNP* with the fork transcription bubble DNA scaffold (no RNA oligo in the scaffold). The initial screen was performed by a sitting-drop vapor diffusion technique. Crystals grown from optimized reservoir solution A (1 μL 0.2 M NaAc, 0.1 M sodium citrate (pH 5.5), and 10% PEG4000 mixed with 1 μL 7.5 mg/mL protein complex) for 3 days at 22°C were harvested for X-ray diffraction data collection. Crystals were soaked in stepwise fashion to reservoir solution A containing 18% (v/v) (2R, 3R)-(-)-2,3-butanediol (Sigma-Aldrich) and cooled in liquid nitrogen. The crystals of σ^H -RNAP derivative containing SeMet-labeled σ^H were obtained by analogous procedure.

Data were collected at Shanghai Synchrotron Radiation Facility (SSRF) beamlines 17U and 19U1, processed using HKL2000⁵⁷. The structure was solved by molecular replacement with Phaser MR⁵⁸ using the structure of *M. smegmatis* core enzyme in a *M. smegmatis* transcription initiation complex (PDB: 5TW1)³⁵ [<https://www.rcsb.org/structure/5TW1>] as the search model. Only one molecule of RNAP core enzyme was found in one asymmetric unit. The electron density maps show clear signal for σ^H . Cycles of iterative model building and refinement were performed in Coot⁵⁹ and Phenix⁶⁰. Residues of σ^H were built into the model at the last stage of refinement. No density of nucleic acid was observed in all stages of refinements, suggesting that the nucleic acids dissociated during crystallization resulting in a crystal of σ^H -RNAP holoenzyme. The final model of *Mtb* σ^H -RNAP holoenzyme was refined to R_{work} and R_{free} of 0.218 and 0.258, respectively. Analogous procedures were used to refine the structures of σ^H -RNAP holoenzyme with SeMet-labeled σ^H .

Structure determination of *M. tuberculosis* σ^H -RNPo. The initial screen of σ^H -RNPo was performed by a sitting-drop vapor diffusion technique. Crystals grown from reservoir solution B (1 μL 2% Tacimate pH 5.0, 0.1 M Sodium citrate pH 5.6, 16% PEG3350 mixed with 1 μL 7.5 mg/mL protein complex) for 15 days at 22°C were harvested for X-ray diffraction data collection. Crystals were soaked in stepwise fashion to the reservoir solution B containing 18% (v/v) (2R, 3R)-(-)-2,3-butanediol (Sigma-Aldrich) and cooled in liquid nitrogen. Data were collected at SSRF beamlines 17U and 19U1, processed using HKL2000⁵⁷. The structure was solved by molecular replacement with Phaser MR⁵⁸ using the structure of *M. tuberculosis* σ^H -RNAP as a search model. Only one molecule of σ^H -RNAP was found in one asymmetric unit. The electron density map showed clear signals for nucleotides in transcription bubble and downstream DNA duplex after initial rigid-body refinement, and clear signals for nucleotides in upstream DNA duplex after iterative

cycles of model building and refinements in Coot⁵⁹ and Phenix⁶⁰. The nucleotides were built into the model at the last stage, and the final model of *Mtb* σ^H -RNPo was refined to R_{work} and R_{free} of 0.220 and 0.255, respectively.

In vitro transcription assay. Transcription assays with *M. tuberculosis* RNAP σ^H -holoenzyme were performed as follows: reaction mixtures contained (20 μL): 80 nM *M. tuberculosis* RNAP core enzyme; 1 μM *M. tuberculosis* σ^H ; 40 mM Tris-HCl, pH 7.9; 75 mM KCl; 5 mM MgCl₂; 2.5 mM DTT; and 12.5% glycerol. Reaction mixtures were incubated for 10 min at 37°C , and then supplemented with 2 μL promoter DNA (1 μM ; amplified from pEASY-Blunt-pCIPB; Supplementary Data 1), and further incubated for 10 min at 37°C . The reaction was initiated by adding 0.7 μL NTP mixture (3 mM [α -³²P]UTP (0.04 Bq/fmol), 3 mM ATP, 3 mM GTP, and 3 mM CTP), and RNA synthesis was allowed to proceed for 10 min at 37°C . Reactions were terminated by adding 8 μL loading buffer (10 mM EDTA, 0.02% bromophenol blue, 0.02% xylene cyanol, and 98% formamide), boiled for 2 min, and stored in ice for 5 min. Reaction mixtures were applied to 15% urea-polyacrylamide slab gels (19:1 acrylamide/bisacrylamide), electrophoresed in 90 mM Tris-borate (pH 8.0) and 0.2 mM EDTA, and analyzed by storage-phosphor scanning (Typhoon; GE Healthcare, Inc.).

Transcription assays with *M. tuberculosis* RNAP σ^A -holoenzyme were performed essentially as above except that σ^A instead of σ^H were added and N25 promoter DNA were used (amplified from pARTaq-N25-100-TR2; Supplementary Table 1 and Supplementary Data 1).

Transcription assays using *M. tuberculosis* RNAP σ^H -holoenzyme and pRv2466c promoters were also performed essentially as above with subtle modifications. The reaction mixtures (20 μL) containing 160 nM *M. tuberculosis* RNAP core enzyme, 1 μM *M. tuberculosis* σ^H , 40 mM Tris-HCl, pH 7.9, 75 mM KCl, 5 mM MgCl₂, 2.5 mM DTT, and 12.5% glycerol were incubated for 10 min at 37°C , and then supplemented with 2 μL promoter DNA (1 μM , amplified from pEASY-pRv2466c; Supplementary Table 1 and Supplementary Data 1), and further incubated for 10 min at 37°C . The reactions were initiated by adding 4 μL NTP mixture (0.1 mM ATP, 0.1 mM GTP, 0.1 mM CTP, and 7 μM [α -³²P]UTP (5.6 Bq/fmol)) and were allowed to proceed for 20 min at 37°C . The reactions were terminated and the transcripts were separated and visualized as above.

Quantification and statistical analysis. All biochemical assays were performed at least three times independently. Data were analyzed with SigmaPlot 10.0 (Systat Software Inc.).

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

Data availability

The accession numbers for the coordinates and structure factors for *M. tuberculosis* σ^H -RNAP and σ^H -RNPo in this paper are PDB: 5ZX3 and 5ZX2, respectively. The source data underlying Figs. 2b, d, 4c, d, f, and 5f, and Supplementary Figs. 2i, 4e, and 5b are provided as a Source Data file. Other data are available from the corresponding author upon reasonable request.

Received: 17 October 2018 Accepted: 1 February 2019

Published online: 11 March 2019

References

- Browning, D. F. & Busby, S. J. Local and global regulation of transcription initiation in bacteria. *Nat. Rev. Microbiol.* **14**, 638–650 (2016).
- Lee, D. J., Minchin, S. D. & Busby, S. J. Activating transcription in bacteria. *Annu. Rev. Microbiol.* **66**, 125–152 (2012).
- Feklistov, A. & Darst, S. A. Structural basis for promoter-10 element recognition by the bacterial RNA polymerase sigma subunit. *Cell* **147**, 1257–1269 (2011).
- Feklistov, A., Sharon, B. D., Darst, S. A. & Gross, C. A. Bacterial sigma factors: a historical, structural, and genomic perspective. *Annu. Rev. Microbiol.* **68**, 357–376 (2014).
- Ruff, E. F., Record, M. T. Jr. & Artsimovitch, I. Initial events in bacterial transcription initiation. *Biomolecules* **5**, 1035–1062 (2015).
- Saecker, R. M., Record, M. T. Jr. & Dehaseth, P. L. 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 (2011).
- Paget, M. S. Bacterial sigma factors and anti-sigma factors: structure, function and distribution. *Biomolecules* **5**, 1245–1265 (2015).
- Gruber, T. M. & Gross, C. A. Multiple sigma subunits and the partitioning of bacterial transcription space. *Annu. Rev. Microbiol.* **57**, 441–466 (2003).

9. Mascher, T. Signaling diversity and evolution of extracytoplasmic function (ECF) sigma factors. *Curr. Opin. Microbiol.* **16**, 148–155 (2013).
10. Staron, A. et al. The third pillar of bacterial signal transduction: classification of the extracytoplasmic function (ECF) sigma factor protein family. *Mol. Microbiol.* **74**, 557–581 (2009).
11. Helmann, J. D. Bacillus subtilis extracytoplasmic function (ECF) sigma factors and defense of the cell envelope. *Curr. Opin. Microbiol.* **30**, 122–132 (2016).
12. Fiebig, A., Herrou, J., Willett, J. & Crosson, S. General stress signaling in the alphaproteobacteria. *Annu. Rev. Genet.* **49**, 603–625 (2015).
13. Woods, E. C. & McBride, S. M. Regulation of antimicrobial resistance by extracytoplasmic function (ECF) sigma factors. *Microbes Infect.* **19**, 238–248 (2017).
14. Flentie, K., Garner, A. L. & Stallings, C. L. *Mycobacterium tuberculosis* transcription machinery: ready to respond to host attacks. *J. Bacteriol.* **198**, 1360–1373 (2016).
15. Rodrigue, S., Proveddi, R., Jacques, P. E., Gaudreau, L. & Manganeli, R. The sigma factors of *Mycobacterium tuberculosis*. *FEMS Microbiol. Rev.* **30**, 926–941 (2006).
16. Sachdeva, P., Misra, R., Tyagi, A. K. & Singh, Y. The sigma factors of *Mycobacterium tuberculosis*: regulation of the regulators. *FEBS J.* **277**, 605–626 (2010).
17. Shultzaberger, R. K., Chen, Z., Lewis, K. A. & Schneider, T. D. Anatomy of *Escherichia coli* sigma70 promoters. *Nucleic Acids Res.* **35**, 771–788 (2007).
18. Yuzenkova, Y., Tadigotla, V. R., Severinov, K. & Zenkin, N. A new basal promoter element recognized by RNA polymerase core enzyme. *EMBO J.* **30**, 3766–3775 (2011).
19. Barne, K. A., Bown, J. A., Busby, S. J. & Minchin, S. D. Region 2.5 of the *Escherichia coli* RNA polymerase sigma70 subunit is responsible for the recognition of the ‘extended-10’ motif at promoters. *EMBO J.* **16**, 4034–4040 (1997).
20. Feklistov, A. et al. A basal promoter element recognized by free RNA polymerase sigma subunit determines promoter recognition by RNA polymerase holoenzyme. *Mol. Cell* **23**, 97–107 (2006).
21. Campbell, E. A. et al. Structure of the bacterial RNA polymerase promoter specificity sigma subunit. *Mol. Cell* **9**, 527–539 (2002).
22. Bae, B., Feklistov, A., Lass-Napiorkowska, A., Landick, R., Darst, S. A. Structure of a bacterial RNA polymerase holoenzyme open promoter complex. *ELife* **4**, e08504 (2015).
23. Feng, Y., Zhang, Y. & Ebright, R. H. Structural basis of transcription activation. *Science* **352**, 1330–1333 (2016).
24. Zhang, Y. et al. Structural basis of transcription initiation. *Science* **338**, 1076–1080 (2012).
25. Koo, B. M., Rhodius, V. A., Nonaka, G., deHaseth, P. L. & Gross, C. A. Reduced capacity of alternative sigmas to melt promoters ensures stringent promoter recognition. *Genes Dev.* **23**, 2426–2436 (2009).
26. Basu, R. S. et al. Structural basis of transcription initiation by bacterial RNA polymerase holoenzyme. *J. Biol. Chem.* **289**, 24549–24559 (2014).
27. Zuo, Y. & Steitz, T. A. Crystal structures of the *E. coli* transcription initiation complexes with a complete bubble. *Mol. Cell* **58**, 534–540 (2015).
28. Duchi, D. et al. RNA polymerase pausing during initial transcription. *Mol. Cell* **63**, 939–950 (2016).
29. Petushkov, I. et al. Interplay between sigma region 3.2 and secondary channel factors during promoter escape by bacterial RNA polymerase. *Biochem. J.* **474**, 4053–4064 (2017).
30. Samanta, S. & Martin, C. T. Insights into the mechanism of initial transcription in *Escherichia coli* RNA polymerase. *J. Biol. Chem.* **288**, 31993–32003 (2013).
31. Rhodius, V. A. et al. Design of orthogonal genetic switches based on a crosstalk map of sigmas, anti-sigmas, and promoters. *Mol. Syst. Biol.* **9**, 702 (2013).
32. Campagne, S., Allain, F. H. & Vorholt, J. A. Extra cytoplasmic function sigma factors, recent structural insights into promoter recognition and regulation. *Curr. Opin. Struct. Biol.* **30**, 71–78 (2015).
33. Lane, W. J. & Darst, S. A. The structural basis for promoter –35 element recognition by the group IV sigma factors. *PLoS Biol.* **4**, e269 (2006).
34. Campagne, S., Marsh, M. E., Capitani, G., Vorholt, J. A. & Allain, F. H. Structural basis for –10 promoter element melting by environmentally induced sigma factors. *Nat. Struct. Mol. Biol.* **21**, 269–276 (2014).
35. Hubin, E. A. et al. Structure and function of the mycobacterial transcription initiation complex with the essential regulator RbpA. *ELife* **6**, e22520 (2017).
36. Lin, W. et al. Structural basis of *Mycobacterium tuberculosis* transcription and transcription inhibition. *Mol. Cell* **66**, 169–179 e168 (2017).
37. Degen, D. et al. Transcription inhibition by the depsipeptide antibiotic salinamide A. *ELife* **3**, e02451 (2014).
38. Manganeli, R. et al. Role of the extracytoplasmic-function sigma factor sigma (H) in *Mycobacterium tuberculosis* global gene expression. *Mol. Microbiol.* **45**, 365–374 (2002).
39. Chakraborty, A. et al. Opening and closing of the bacterial RNA polymerase clamp. *Science* **337**, 591–595 (2012).
40. Raman, S. et al. The alternative sigma factor SigH regulates major components of oxidative and heat stress responses in *Mycobacterium tuberculosis*. *J. Bacteriol.* **183**, 6119–6125 (2001).
41. Song, T., Song, S. E., Raman, S., Anaya, M. & Husson, R. N. Critical role of a single position in the –35 element for promoter recognition by *Mycobacterium tuberculosis* SigE and SigH. *J. Bacteriol.* **190**, 2227–2230 (2008).
42. Kim, K. Y., Park, J. K. & Park, S. In *Streptomyces coelicolor* SigR, methionine at the –35 element interacting region 4 confers the –31’-adenine base selectivity. *Biochem. Biophys. Res. Commun.* **470**, 257–262 (2016).
43. Juang, Y. L. & Helmann, J. D. 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 (1994).
44. Fenton, M. S., Lee, S. J. & Gralla, J. D. *Escherichia coli* promoter opening and –10 recognition: mutational analysis of sigma70. *EMBO J.* **19**, 1130–1137 (2000).
45. Campbell, E. A. et al. Crystal structure of *Escherichia coli* sigmaE with the cytoplasmic domain of its anti-sigma RseA. *Mol. Cell* **11**, 1067–1078 (2003).
46. Sainsbury, S., Niesser, J. & Cramer, P. Structure and function of the initially transcribing RNA polymerase II-TFIIB complex. *Nature* **493**, 437–440 (2013).
47. Murakami, K. S., Masuda, S. & Darst, S. A. Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. *Science* **296**, 1280–1284 (2002).
48. Kulbachinskiy, A. & Mustaev, A. 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 (2006).
49. Pupov, D., Kuzin, I., Bass, I. & Kulbachinskiy, A. Distinct functions of the RNA polymerase sigma subunit region 3.2 in RNA priming and promoter escape. *Nucleic Acids Res.* **42**, 4494–4504 (2014).
50. Lerner, E. et al. Backtracked and paused transcription initiation intermediate of *Escherichia coli* RNA polymerase. *Proc. Natl Acad. Sci. USA* **113**, E6562–E6571 (2016).
51. Liu, X., Bushnell, D. A. & Kornberg, R. D. Lock and key to transcription: sigma-DNA interaction. *Cell* **147**, 1218–1219 (2011).
52. Vvedenskaya, I. O. et al. Massively systematic transcript end readout, “MASTER”: transcription start site selection, transcriptional slippage, and transcript yields. *Mol. Cell* **60**, 953–965 (2015).
53. Heyduk, E., Kuznedelov, K., Severinov, K. & Heyduk, T. A consensus adenine at position –11 of the nontemplate strand of bacterial promoter is important for nucleation of promoter melting. *J. Biol. Chem.* **281**, 12362–12369 (2006).
54. Lim, H. M., Lee, H. J., Roy, S. & Adhya, S. A. “master” in base unpairing during isomerization of a promoter upon RNA polymerase binding. *Proc. Natl Acad. Sci. USA* **98**, 14849–14852 (2001).
55. Tomsic, M. et al. Different roles for basic and aromatic amino acids in conserved region 2 of *Escherichia coli* sigma(70) in the nucleation and maintenance of the single-stranded DNA bubble in open RNA polymerase-promoter complexes. *J. Biol. Chem.* **276**, 31891–31896 (2001).
56. Banerjee, R., Rudra, P., Prajapati, R. K., Sengupta, S. & Mukhopadhyay, J. Optimization of recombinant *Mycobacterium tuberculosis* RNA polymerase expression and purification. *Tuberculosis (Edinb.)* **94**, 397–404 (2014).
57. Otwinowski, Z. & Minor, W. [20] Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
58. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
59. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
60. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).

Acknowledgements

Work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB29020000), the National Natural Science Foundation of China (31670067 and 31822001), and the Leading Science Key Research Program of the Chinese Academy of Sciences (QYZDB-SSW-SMC005). We thank Prof. Richard Ebright for generous gifts of pACYCDuet-Mtb-rpoA-rpoD, pETDuet-Mtb-rpoB-rpoC, and pARTaq-N25-100-TR2 constructs; 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.

Author contributions

L.L. solved the structures. C.F. performed in vitro transcription assays. N.Z. assisted in structure determination. T.W. prepared pACYCDuet-*Mtb-rpoA-rpoZ* and purified proteins. Y.Z. designed experiments, analyzed data, and wrote the manuscript.

Additional information

Supplementary Information accompanies this paper at <https://doi.org/10.1038/s41467-019-09096-y>.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

Journal peer review information: *Nature Communications* thanks Xiaodong Zhang and the other anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

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) 2019