Structural basis of ECF-σ-factor-dependent transcription initiation

Wei Lin1,3, Sukhendu Mandal1, David Degen1, Min Sung Cho1, Yu Feng1, Kalyan Das2 & Richard H. Ebright1

Extracytoplasmic (ECF) σ factors, the largest class of alternative σ factors, are related to primary σ factors, but have simpler structures, comprising only two of six conserved functional modules in primary σ factors: region 2 (σR2) and region 4 (σR4). Here, we report crystal structures of transcription initiation complexes containing Mycobacterium tuberculosis RNA polymerase (RNAP), M. tuberculosis ECF σ factor σL, and promoter DNA. The structures show that σR2 and σR4 of the ECF σ factor occupy the same sites on RNAP as in primary σ factors, show that the connector between σR2 and σR4 of the ECF σ factor—although shorter and unrelated in sequence—follows the same path through RNAP as in primary σ factors, and show that the ECF σ factor uses the same strategy to bind and unwind promoter DNA as primary σ factors. The results define protein-protein and protein-DNA interactions involved in ECF-σ-factor-dependent transcription initiation.
Bacterial transcription initiation is carried out by an RNA polymerase (RNAP) holoenzyme comprising RNAP core enzyme and a σ factor. Bacteria contain a primary σ factor (group-1 σ factor; \( \sigma^{E} \) in *Escherichia coli*; \( \sigma^{k} \) in other bacteria) that mediates transcription initiation at most genes required for growth under most conditions and sets of alternative σ factors that mediate transcription initiation at sets of genes required in certain cell types, developmental states, or environmental conditions.

Group-1 σ factors contain six conserved functional modules: \( \sigma \) regions 1.1, 1.2, 2, 3, 3/4 linker, and 4 (\( \sigma R1.1, \sigma R1.2, \sigma R2, \sigma R3, \sigma R3/4, \sigma R4 \); Fig. 1a). \( \sigma R1.1 \) plays a regulatory role, inhibiting interactions between free, non-RNAP-bound, \( \sigma \) and DNA. \( \sigma R1.2, \sigma R2, \sigma R3, \) and \( \sigma R4 \) play roles in promoter recognition. \( \sigma R2 \) and \( \sigma R4 \) recognize the promoter -10 element and the promoter -35 element, respectively, and \( \sigma R1.2 \) and \( \sigma R3 \) recognize sequences immediately downstream and immediately upstream, respectively, of the promoter -10 element. The \( \sigma R3/4 \) linker plays multiple crucial roles. The \( \sigma R3/4 \) linker connects \( \sigma R2 \) to \( \sigma R4 \); the \( \sigma R3/4 \) linker enters the RNAP active-center cleft, where it interacts with template-strand ssDNA of the unwound transcription bubble, pre-organizing template-strand ssDNA to adopt a helical conformation and to engage the RNAP active center, thereby facilitating initiating-nucleotide binding and de novo transcription initiation; and the \( \sigma R3/4 \) linker exits the RNAP active-center cleft by threading through the RNAP RNA exit channel.

Before RNA synthesis takes place, the \( \sigma R3/4 \) linker serves as a molecular mimic of RNA, or molecular placeholder for RNA, through its interactions with template-strand ssDNA and the RNAP RNA exit channel. As RNA synthesis takes place, the \( \sigma R3/4 \) linker then is displaced—off of template-strand ssDNA and out of the RNAP RNA exit channel—driven by steric interactions with the 5'-end of the nascent RNA. The \( \sigma R3/4 \) linker must be displaced from template-strand ssDNA during initial transcription; this requirement imposes energy barriers associated with initial-transcription pausing and abortive initiation. The \( \sigma R3/4 \) linker must be displaced from the RNAP RNA exit channel during the transition between initial transcription and transcription elongation; this requirement imposes energy barriers that are exploited to trigger promoter escape and to transform the transcription initiation complex into the transcription elongation complex.

Crystal structures of RNAP holoenzyme and transcription initiation complexes containing group-1 σ factors define the protein–protein and protein–nucleic acid interactions involved in group-1-σ-factor-dependent transcription initiation, and extensive biochemical and biophysical characterization defines the protein–protein and protein–nucleic acid interactions and mechanisms involved in group-1-σ-factor-dependent transcription initiation.2,3,6,7,12-19

Alternative σ factors—with the exception of the alternative σ factor mediating the response to nitrogen starvation (\( \sigma^{N} \) in *E. coli*; \( \sigma^{N} \) in other bacteria)20,21—are members of the same protein family as group-1 σ factors. Group-2 and group-3 alternative σ factors are closely related in structure to group-1 σ factors, lacking only functional modules \( \sigma R1.1 \) (in group-2 σ factors) or \( \sigma R1.1 \) and \( \sigma R1.2 \) (in group-3 σ factors). The close structural similarity of group-2 and group-3 σ factors to group-1 σ factors, together with crystal structures of transcription initiation complexes containing group-2 σ factors22, facilitates an understanding of the mechanism of group-2- and group-3-σ-factor-dependent transcription initiation.

Group-4 alternative σ factors—also referred to as “extracytoplasmic σ factors” (ECF σ factors), based on functional roles in response to cell-surface and other extracytoplasmic stresses—are only distantly related to group-1 σ factors and are substantially smaller than group-1 σ factors, lacking four of the six functional modules present in group-1 σ factors (Fig. 1a).13-30 ECF σ factors comprise only a module related to \( \sigma R2 \) (the module that recognizes promoter -10 elements in group-1 σ factors), a module related to \( \sigma R4 \) (the module that recognizes promoter -35 elements in group-1 σ factors), and a short \( \sigma R2/4 \) linker that has no detectable sequence similarity to the \( \sigma R3/4 \) linker of group-1 σ factors. No structural information previously has been reported for RNAP holoenzymes or transcription initiation complexes containing ECF σ factors. In the absence of structural information for ECF σ factors, it has been unclear how ECF σ factors, despite lacking sequences homologous to the \( \sigma R3/4 \) linker of group-1 σ factors, are able to connect \( \sigma R2 \) and \( \sigma R4 \) with an appropriate spacing to recognize promoter -10 and -35 elements, are able to pre-organize the DNA template strand to facilitate initiating-nucleotide binding and de novo transcription initiation; and are able to coordinate entry of RNA into the RNA-exit channel with promoter escape. In addition, in the absence of structural information, and with comparatively limited sequence similarity between \( \sigma R2 \) of ECF σ factors and \( \sigma R2 \) of group-1 σ factors,13-15 it has been unclear whether \( \sigma R2 \) of ECF σ factors adopts the same fold as \( \sigma R2 \) of group-1 σ factors and uses the same strategy to bind and unwind the promoter -10 element as group-1 σ factors.

ECF σ factors are numerically the largest, and functionally the most diverse, alternative σ factors.1,24-30 Fully 10 of the 13 σ factors in *Mycobacterium tuberculosis* (Mtb), the causative agent of tuberculosis, are ECF σ factors: \( \sigma^{E} \), \( \sigma^{K} \), \( \sigma^{E} \), \( \sigma^{K} \), \( \sigma^{L} \), \( \sigma^{E} \), \( \sigma^{K} \), \( \sigma^{M} \), \( \sigma^{M} \), \( \sigma^{M} \), \( \sigma^{M} \), \( \sigma^{M} \), \( \sigma^{M} \) and \( \sigma^{M} \), mediating responses to nutrition depletion, surface stress, temperature stress, oxidative stress, pH stress, growth in stationary phase, and growth in macrophages. For example, the Mtb ECF σ factor \( \sigma^{E} \) (Supplementary Fig. 1A) mediates the response to oxidative stress and regulates its own synthesis, polynucleotide synthase synthesis, cell-wall synthesis, lipid transport, the oxidative state of exported proteins, and virulence.36-38

In this work, we have determined crystal structures, at 3.3-3.8 Å resolution, of functional transcription initiation complexes comprising Mtb RNAP, the Mtb RNAP ECF σ factor \( \sigma^{E} \), and nucleic-acid scaffolds corresponding to the transcription bubble and downstream dsDNA of an ECF-σ-factor-dependent RNAP-promoter open complex (Mtb RPo-σ\(^{E}\)) or an RNAP-promoter initial transcribing complex (Mtb RPitc-σ\(^{E}\)) (Table 1; Fig. 1; Supplementary Figs. 1, 2).

**Results**

**Structures of Mtb RPo-σ\(^{E}\) and Mtb RPitc-σ\(^{E}\).** Structures were determined using recombinant Mtb RNAP core enzyme prepared by co-expression of Mtb RNAP subunit genes in *E. coli*, recombinant Mtb \( \sigma^{L} \), and synthetic nucleic-acid scaffolds based on the sequence of the \( \sigma^{E} \)-dependent promoter P-sigL. (the promoter responsible for expression of the gene encoding \( \sigma^{E} \))36-38 (Supplementary Figs. 1, 2). Transcription experiments demonstrate that Mtb RNAP-σ\(^{E}\) holoenzyme (containing the group-1 σ factor \( \sigma^{E}\)) does not efficiently perform transcription initiation at the P-sigL promoter, whereas Mtb RNAP-σ\(^{E}\) holoenzyme (containing the ECF σ factor \( \sigma^{E}\)) does (Supplementary Fig. 1E). We prepared “downstream-fork-junction” nucleic-acid scaffolds containing P-sigL sequences, analogous to the downstream-fork-junction nucleic-acid scaffolds containing consensus group-1-σ-factor-dependent promoter sequences used previously for structural analysis of group-1-σ-factor-dependent transcription initiation (Supplementary Fig. 2, left panels). Because the P-sigL transcription start site (TSS) had been mapped only provisionally,36-38 we prepared and analyzed a set of downstream-fork-junction nucleic-acid scaffolds having different lengths—4 nt, 5 nt, 6, or 7 nt—of
the “spacer” between the P-sigL promoter -10 region and down-stream dsDNA (Supplementary Fig. 2, left panels). Transcription experiments indicated that all analyzed nucleic-acid scaffolds were functional in σL-dependent de novo transcription initiation at the expected TSS (with the initiating nucleotide base-pairing to template-strand ssDNA 2 nt upstream of dsDNA), and σL-dependent primer-dependent transcription initiation at the expected TSS (with the primer 3′ nucleotide base-pairing to template-strand ssDNA 2 nt upstream of dsDNA), with highest levels of function observed for a spacer length of 6 nt (Supplementary Fig. 1F, G). Robotic crystallization trials identified crystallization conditions yielding high-quality crystals for spacer lengths of 4 nt, 5 nt, or 6 nt (Table 1; Supplementary Fig. 2, center panels). X-ray datasets were collected at synchrotron beam sources, and structures were solved by molecular replacement and refined to 3.3–3.8 Å resolution (Table 1; Supplementary Fig. 2, right panels). Experimental electron-density maps showed clear density for RNAP, σL, and nucleic acids (Supplementary Fig. 2, right panels). The resulting structures were essentially identical for nucleic-acid scaffolds having spacer lengths of 4 nt, 5 nt, or 6 nt (Supplementary Fig. 2, right panels). However, map quality was highest for the nucleic-acid scaffold having a spacer length of 6 nt, and therefore subsequent analysis focussed on structures with a spacer length of 6 nt (Mtb RPitc5-σL_sp6). For the nucleic-acid scaffold containing a 6 nt spacer, the translocational state of the transcription complex was experimentally verified by preparation of a scaffold having a single 5-bromo-dU substitution and collection of bromine anomalous diffraction data (Table 1; Supplementary Fig. 2D). The fit of σL separately was experimentally verified by preparation of a selenomethionine-labeled σL derivative and collection of selenium anomalous diffraction data (Table 1; Supplementary Fig. 2E).
| Structure | Mtb RpITcS-σi_sp4 | Mtb RpITcS-σi_sp5 | Mtb RpITcS-σi_sp6 |
|-----------|------------------|------------------|------------------|
| PDB code  | 6DV9             | 6DVB             | 6DVC             |
| **Data collection** |                   |                  |                  |
| Source    | APS 19-ID        | SSRL-9-2         | APS 19-ID        |
| Space group| P2₁,2₁           | P2₁,2₁           | P2₁,2₁           |
| Cell dimensions | a, b, c (Å) | a, b, c (Å) | a, b, c (Å) |
| α, β, γ (°) | 143.3, 161.4, 237.7 | 143.7, 160.6, 240.4 | 146.3, 161.5, 240.6 |
| Resolution (Å) | 50.0-3.8 (3.9-3.8) | 50.0-3.8 (3.9-3.8) | 50.0-3.3 (3.4-3.3) |
| Number of unique reflections | 53,020 | 53,728 | 98,083 |
| Rmerge | 0.162 (0.567) | 0.188 (0.706) | 0.175 (0.710) |
| Rfree | 0.172 (0.613) | 0.200 (0.752) | 0.184 (0.764) |
| CC1/2 (highest resolution shell) | 0.526 | 0.715 | 0.558 |
| Completeness (%) | 96.6 (90.5) | 97.1 (96.5) | 98.9 (99.3) |
| Redundancy | 10.7 (9.9) | 8.2 (8.1) | 10.4 (7.4) |
| Anomalous completeness (%) | N/A | N/A | N/A |
| Anomalous redundancy | N/A | N/A | N/A |
| **Refinement** |                   |                  |                  |
| Resolution (Å) | 50.0-3.8 | 50.0-3.8 | 50.0-3.3 |
| Number of unique reflections | 52,512 | 53,617 | 85,687 |
| Number of test reflections | 2625 | 2691 | 4267 |
| Rwork/Rfree | 0.18/0.23 (0.29/0.32) | 0.20/0.24 (0.32/0.35) | 0.19/0.23 (0.34/0.35) |
| Number of atoms | Protein: 24,956 | 24,974 | 24,982 |
| | Ligand/ion: 3 | 2 | 2 |
| r.m.s. deviations | Bond lengths (Å): 0.002 | 0.002 | 0.002 |
| | Bond angles (°): 0.489 | 0.457 | 0.468 |
| MolProbity statistics | Clash score: 7.2 | 6.3 | 6.0 |
| | Rotamer outliers (%): 1.6 | 2.4 | 2.6 |
| | Cβ outliers (%): 0 | 0 | 0 |
| Ramachandran plot | Favored (%): 95.2 | 95.4 | 95.5 |
| | Outliers (%): 0.3 | 0.2 | 0.3 |

aData for the highest resolution shell are presented in parentheses
bRmerge values for 6DVB, 6DVC, 6DVD, and 6DVE reflect an anisotropic component
Interactions between ECF σ factor and RNAP. The structural organization of the ECF σ\textsuperscript{L}-factor-dependent transcription initiation complex is unexpectedly similar to that of a group-1 σ\textsuperscript{A}-factor-dependent transcription initiation complex (Figs. 1b and 2). σR2 and σR4 of σ\textsuperscript{L} occupy the same positions on RNAP, and make the same interactions with RNAP, as σR2 and σR4 of σ\textsuperscript{A} factor (Fig. 2). Despite the smaller size of the connector between σR2 and σR4 in σ\textsuperscript{L} as compared to σ\textsuperscript{A} (20 residues vs. 84 residues if one includes σR3; 20 residues vs. 28 residues if one does not include σR3; Supplementary Fig. 1A), the connector in σ\textsuperscript{L} spans the full distance between the σR2 and σR4 binding positions on RNAP and follows a path through RNAP similar to that of the connector in σ\textsuperscript{A} (Fig. 2). Thus, the σ\textsuperscript{L}-σR2/4 linker, like the σ\textsuperscript{A} σR3/4 linker\textsuperscript{2-4,6,7,12-19}, first enters the RNAP active-center cleft and approaches the RNAP active center, and then makes a sharp turn and exits the RNAP active-center cleft through the RNAP RNA-exit channel.

Inside the RNAP active-center cleft, the σ\textsuperscript{L}-σR2/4 linker, like the σ\textsuperscript{A} σR3/4 linker\textsuperscript{6,7,14-19}, makes direct interactions with template-strand ssDNA nucleotides of the unwound transcription bubble (Figs. 2–4, Supplementary Fig. 3A). The interactions of the σ\textsuperscript{L}-σR2/4 linker with template-strand ssDNA include a direct H-bonded interaction of σ\textsuperscript{L}-Ser96 with a Watson–Crick H-bonding atom of the template-strand nucleotide at promoter position −5 (Fig. 3b; Supplementary Fig. 3A, bottom). The interactions of the σ\textsuperscript{L}-σR2/4 linker with template-strand ssDNA are similar to, but less extensive than, those of the σ\textsuperscript{A} σR3/4 linker with template-strand ssDNA, which include direct H-bonded interactions of σ\textsuperscript{A} Asp432 and Ser433 with Watson–Crick H-bonding atoms of template-strand ssDNA nucleotides at promoter positions −4 and −3 (Fig. 3a; Supplementary Fig. 3A).

In the case of the group-1 σ factor, σ\textsuperscript{A}, the interactions between this segment of the σR3/4 linker and template-strand ssDNA pre-organize template-strand ssDNA and facilitate initiating-nucleotide binding and de novo initiation.

In the case of the group-1 σ factor, σ\textsuperscript{A}, the interactions between this segment of the σR3/4 linker and template-strand ssDNA must be broken, and this segment of the σR3/4 linker must be displaced, when the nascent RNA reaches a length >4 nt during initial transcription, and this requirement for breakage of interactions and displacement is thought to impose an energy barrier that results in, or enhances, abortive initiation\textsuperscript{2,5,6,7,8} and initial-transcription pausing\textsuperscript{9-11}. The similarity of the interactions made by the ECF σ factor, σ\textsuperscript{L}, suggests that ECF σ factors likewise have a similar requirement for displacement of a linker segment during initial transcription—in this case, when the nascent RNA reaches a length of >5 nt (Supplementary Fig. 3B)—and that this similar requirement imposes an energy barrier that results in, or enhances, abortive initiation and initial-transcription pausing.

Consistent with this hypothesis, transcription and transcript-release experiments indicate that Mtb RNAP-σ\textsuperscript{L} holoenzyme efficiently performs abortive initiation, producing and releasing short abortive RNA products (Supplementary Fig. 3C).

The five C-terminal residues of the σ\textsuperscript{L}-σR2/4 linker, like the ten C-terminal residues of the σ\textsuperscript{A} σR3/4 linker, exit the RNAP active-center cleft and connect to σR4 by threading through the RNAP RNA-exit channel (Fig. 2). In the case of the group-1 σ factor, σ\textsuperscript{A}, the C-terminal segment of the σR3/4 linker must be displaced from the RNA-exit channel when the nascent RNA reaches a length of 11 nt at the end of initial transcription and moves into the RNA-exit channel, and this displacement is thought to alter interactions between σR4 and RNAP, thereby triggering promoter escape and transforming the transcription initiation complex into a transcription elongation complex\textsuperscript{2-4}. The similarity of the threading through the RNAP RNA-exit channel by the ECF σ factor, σ\textsuperscript{L}, suggests that ECF σ factors have a similar requirement for displacement of a linker C-terminal segment and have a similar mechanism of promoter escape and transformation from transcription initiation complexes into transcription elongation complexes.

In its interactions with template-strand ssDNA and the RNAP RNA-exit channel, the σ\textsuperscript{L}-σR2/4 linker, like the σ\textsuperscript{A} σR3/4 linker...
**Fig. 3** Protein–nucleic acid interactions by group-1 and ECF σ factors: summary. 

**a** Summary of protein–nucleic acid interactions in *Mtb* RPitc-σ<sup>76</sup>. Black residue numbers and lines, interactions by *Mtb* RNAP; green residue numbers and lines, interactions by *Mtb* σ<sup>76</sup>; blue, -10 element of DNA nontemplate strand; light blue, discriminator element of DNA nontemplate strand; pink, rest of DNA nontemplate strand; red, DNA template strand; magenta, RNA product; violet circle, RNAP active-center Mg<sup>2+</sup>; cyan boxes, bases unstacked and inserted into protein pockets. Residues are numbered as in *Mtb* RNAP. 

**b** Summary of protein–nucleic acid interactions in *Mtb* RPitc-σ<sup>3</sup>. Green residue numbers and lines, interactions by *Mtb* σ<sup>3</sup>. Other colors are as in (a). See Supplementary Figs. 3-5 and 7.
σR3/4 linker²⁴, appears to serve as a molecular mimic, or a molecular placeholder, for nascent RNA, making interactions with template-strand ssDNA and the RNAP RNA-exit channel in early stages of transcription initiation that subsequently, in late stages of transcription initiation and in transcription elongation, are made by nascent RNA. The σ⁰ σR2/4 linker and the σ³ σR3/4 linker, both have net negative charge (Supplementary Fig. 1A), and both employ extended conformations (fully extended for the σ³ σR2/4 linker; largely extended for the σ³ σR3/4 linker; Fig. 2) to interact with template-strand ssDNA and the RNAP RNA-exit channel, consistent with function as molecular mimics of a negatively charged, extended nascent RNA. Nevertheless, the σ⁰ σR2/4 linker and the σ³ σR3/4 linker exhibit no detectable sequence similarity (Supplementary Fig. 1A) and no detailed structural similarity (Fig. 2). We conclude that the σR2/4 linker of an ECF σ factor and the σR3/4 linker of a group-1 σ factor provide an example of functional analogy in the absence of structural homology.

Interactions between ECF σ factor and promoter -10 element. The structure reveals the interactions between the ECF σ factor, σ³, and promoter DNA that mediate recognition of the promoter -10 element (Figs. 3–5; Supplementary Fig. 4). The σ³ conserved module σR2, like the σ³ conserved module σR2, mediates recognition of the promoter -10 element through interactions with nontemplate-strand ssDNA in the unwound transcription bubble (Figs. 3 and 4). In the case of the group-1 σ factor, σ³, a crucial aspect of recognition of the promoter -10 element is unstacking of nucleotides, flipping of nucleotides, and insertion of nucleotides into protein pockets at two positions of the σ³-dependent promoter⁶,³⁹; i.e., position -11 (referred to as the “master nucleotide”, based on its especially important role in promoter recognition)⁴⁰ and position -7 (Figs. 3a and 4a; Supplementary Fig. 4). The ECF σ factor, σ³, unstacks nucleotides, flips nucleotides, and inserts nucleotides into protein pockets at the corresponding positions of the σ³-dependent promoter (here designated positions “-11” and “-7”; Figs. 3b, 4b, and 5; Supplementary Fig. 4) and also unstacks and inserts a nucleotide into a protein pocket at one additional position of the σ³-dependent promoter (position “-12”; Figs. 4b and 5).

RNAP σ²-holoenzyme unstacks, flips, and inserts into a protein pocket a guanosine at position “-11” of the σ³-dependent promoter, making extensive interactions with the base moiety of the guanosine, including multiple direct H-bonded interactions with Watson–Crick H-bonding atoms (Figs. 3, 4b, and 5; Supplementary Fig. 4).

**Fig. 4** Protein–nucleic acid interactions by group-1 and ECF σ factors: interactions with transcription bubble. **a** Left: interactions of Mtb RNAP and σ³ with transcription-bubble nontemplate strand, transcription-bubble template strand, and downstream dsDNA. Right: interactions of σ³ σR2 with σ³-dependent promoter -10 element. For promoter positions -11 (“master nucleotide”)⁴⁰ and -7, bases are unstacked and inserted into pockets (cyan boxes). Colors are as in Figs. 1-3. **b** Left: interactions of Mtb RNAP and σ³ with transcription-bubble nontemplate strand, transcription-bubble template strand, and downstream dsDNA. Right: interactions of σ³ σR2 with σ³-dependent promoter -10 element. For two promoter positions, here designated “-11” (“master nucleotide”) and “-7”, by analogy to corresponding nucleotides in group-1-σ-factor complex (panel a), bases are unstacked and inserted into pockets (cyan boxes). For one additional nucleotide, here designated “-12”, the base also appears to be unstacked and inserted into a pocket (dashed cyan box). See Supplementary Figs. 4, 5 and 7.
The interactions between σL and guanosine at position “-11” of the σL-dependent promoter are similar to the interactions between RNAP σA holoenzyme and adenosine at position -11 of the σA-dependent promoter -10 element, including, in particular, similar stacking interactions of σL aromatic amino acid Trp68 with guanosine and of corresponding σA aromatic amino acid Tyr436 with adenosine (Supplementary Fig. 4A). The different specificities—guanosine at position “-11” for σL vs. adenosine at position -11 for σA—arise from differences in the H-bond-donor/H-bond-acceptor character of atoms forming the floors of the relevant protein pockets of σL and σA, with H-bonding complementarity to guanosine in σL and H-bonding complementarity to adenosine in σA (Supplementary Fig. 4A).

RNAP σL holoenzyme also unstacks, flips, and inserts into a protein pocket a guanosine at position “-7” of the σL-dependent promoter, making extensive interactions with the base moiety of the guanosine, including a direct H-bonded interaction with a Watson–Crick H-bonding atom (Figs. 3, 4b, and 5; Supplementary Fig. 4B). These interactions are similar in location to, but different in detail from, the interactions made by RNAP σA holoenzyme with thymidine at position -7 of the σA-dependent promoter (Fig. 4; Supplementary Fig. 4B). The differences in detail arise from the fact that σA does not contain conserved module σR1.2. In the case of σL, the interactions involve residues of σR2 and residues of RNAP β subunit, with the base moiety of the guanosine at position “-7” being inserted into a cleft between σR2 and β (Fig. 5; Supplementary Fig. 4B). In contrast, in the case of σA, the interactions involve residues of σR2 and residues of σR1.2, with the base moiety of the thymidine at position -7 being inserted into a cleft between σR2 and σR1.2 (Supplementary Fig. 4B).

RNAP σL holoenzyme also appears to unstack and insert into a protein pocket a thymidine at position “-12” of the σL-dependent promoter (Figs. 4b and 5), placing one face of the base moiety of the thymidine in a shallow surface pocket, in position to make a direct H-bonded interaction with a Watson–Crick atom (Fig. 5). The interaction with an unstacked nucleotide inserted into a protein pocket implies that position “-12” of the σL-dependent promoter must be ssDNA in RPo-σL and RPot-σL, and thus that the transcription bubble must extend to position “-12” in RPo-σL and RPot-σL. This interaction does not have a counterpart in the σA-dependent transcription initiation complex, in which position -12 of the σA-dependent promoter is dsDNA and in which the transcription bubble extends only to position -11.15–19.

In addition to these potential specificity-determining interactions with unstacked nucleotides inserted into protein pockets, RNAP σL holoenzyme makes potentially specificity-determining interactions with positions “-9” and “-8” of the σL-dependent promoter (Fig. 5). RNAP σL holoenzyme makes a direct H-bonded interaction, through RNAP β subunit, with a Watson–Crick atom of the base moiety of cytidine at position -9. Additionally, RNAP σL holoenzyme makes a direct H-bonded interaction with a Watson–Crick atom of the base moiety of thymidine at position -8.
“-9” (Fig. 5) and makes two direct H-bonded interactions, through σR2 and RNAP β subunit, with a Watson–Crick atom of the base moiety of adenosine at position “-8” (Fig. 5).

Biochemical experiments assessing effects of all possible single-base-pair substitutions at each position of the P-sigL promoter -10 region confirm the functional significance of the positions contacted in the crystal structure (positions “-12”, “-11”, “-9”, “-8”, and “-7”; Fig. 6a), confirm the sequence preferences at these positions inferred from the crystal structure (Fig. 6a), and yield a revised consensus sequence for the σL-dependent -10 element of

### Fig. 6
Recognition by Mtb σL of σL-promoter -10 element: experimental data. 

**a** Systematic-substitution experiments defining σL-dependent promoter -10-element consensus sequence. Relative transcriptional activities of derivatives of σL-dependent promoter P-sigL having all possible single base-pair substitutions at each position of promoter -10 element, “-12” through “-7”. Inferred consensus nucleotides are shown at the bottom, and data for inferred consensus nucleotides are hatched. Error bars, SE (N = 3). **b** Sequence logo for σL-promoter -10-element consensus sequence [generated using transcription data from (a) and enoLOGOS71 (http://biodev.hgen.pitt.edu/enologos/); input setting “energy (2)” and weight-type setting “probabilities”]. **c** Alanine-scanning experiments41 demonstrating functional importance of observed amino acid-base interactions in recognition of σL-promoter -10 element. Effects on transcription of alanine substitutions of σL amino acids that contact σL-dependent promoter -10 element, positions “-12” through “-7” (identities of contacting amino acids from Figs. 3 and 5). **d, e** Loss-of-contact experiments42–45 indicating that σL residues His54 and Asp60 determine specificity at position “-12” and “-11”, respectively. Left: transcriptional activity with wild-type σL for all possible single base-pair-substitutions at indicated position (strong specificity for consensus base pair). Right: transcriptional activity of σL derivatives having alanine substitutions (no specificity for consensus base pair). Error bars, SE (N = 3). See Supplementary Fig. 6. Source data are provided as a Source Data file.
Interactions between ECF σ factor and promoter CRE. The structure reveals the interactions between RNAP σ^I^-holoenzyme and nontemplate-strand ssDNA downstream of the promoter -10 element in the ECF σ^I^-dependent transcription initiation complex (Fig. 3b; Supplementary Fig. S5). In the case of group-1-σ-factor-dependent transcription initiation complexes, sequence-specific interactions occur between RNAP β subunit and a 6 nt segment of nontemplate-strand ssDNA downstream of the promoter -10 element referred to as the “core recognition element” (CRE; positions -6 through +2).5,6,19,46 These interactions include, most notably, (1) stacking of a tryptophan residue of RNAP β subunit on the base moiety of thymidine at nontemplate-strand position +1 (Supplementary Fig. 5A), and (2) unstacking, flipping, and insertion into a protein pocket, formed by the RNAP β subunit, of the guanosine at nontemplate-strand position +2 (Figs. 3 and 4a; Supplementary Fig. 5B). The identical interactions occur in the ECF σ^II^-dependent transcription initiation complex (Figs. 3 and 4b; Supplementary Fig. 5).

Biochemical experiments assessing effects of all possible base-pair substitutions at positions downstream of the P-sigL promoter -10 element (positions -4 through +2) confirm the functional significance of the interactions in the crystal structure with thymidine at position +1 and guanosine at position +2 (Supplementary Fig. 6A) and yield a CRE consensus sequence for an ECF σ^II^-dependent transcription initiation complex (Supplementary Fig. 6B) similar to the CRE consensus sequence for a group-1-σ-factor-dependent transcription initiation complex.6,46 Three of four characterized Mbσ^II^-dependent promoters36,37 match the CRE consensus sequence of Supplementary Fig. 6B at position +2, the position at which a guanosine is unstacked, flipped, and inserted into a protein pocket (P-sigL, P-pks10, and P-Rv1139c).

Discussion

Our structural results show that: (1) σR2 and σR4 of an ECF σ factor σ^II^ adopt the same folds and interact with the same sites on RNAP as σR2 and σR4 of a group-1 σ factor (Figs. 1 and 2); (2) the connector between σR2 and σR4 of ECF σ factor σ^II^ enters the RNAP active-center cleft to interact with template-strand ssDNA and then exits the RNAP active-center cleft by threading through the RNAP RNA-exit channel in a manner functionally analogous—but not structurally homologous—to the connector between σR2 and σR4 of a group-1 σ factor (Figs. 1 and 2; Supplementary Fig. 3); (3) ECF σ factor σ^II^ recognizes the -10 element of a σ^II^-dependent promoter by unstacking nucleotides and inserting nucleotides into protein pockets at three positions of the transcription-bubble nontemplate-strand ssDNA (positions -12, -11, and -7; Figs. 3–5; Supplementary Fig. 4), and (4) RNA polymerase recognizes the CRE of a σ^II^-dependent promoter by stacking a nucleotide on a tryptophan and by unstacking, flipping, and inserting a nucleotide into a protein pocket (positions +1 and +2; Figs. 3 and 4; Supplementary Fig. 5). Our biochemical results confirm the functional significance of the observed protein-DNA interactions with the -10 element and CRE of a σ^II^-dependent promoter (Fig. 6a; Supplementary Fig. 6A), provide consensus sequences for the -10 element and CRE of a σ^II^-dependent promoter (Fig. 6b; Supplementary Fig. 6B), and define individual specificity-determining amino-acid–base interactions for two positions of the -10 element of a σ^II^-dependent promoter (positions -12 and -11; Figs. 6c, d). The results provide an indispensable foundation for understanding the structural and mechanistic basis of ECF-σ-factor-dependent transcription initiation.

Our results regarding the connector between σR2 and σR4 of an ECF σ factor, in conjunction with previous results, indicate that all classes of bacterial σ factors contain structural modules that enter the RNAP active-center cleft to interact with template-strand ssDNA and then leave the RNAP active-center cleft by threading through the RNAP RNA-exit channel, providing mechanisms to facilitate de novo initiation, to coordinate extension of the nascent RNA with abortive initiation and initial-transcription pausing, and to coordinate entry of RNA into the RNA-exit channel with promoter escape. For ECF σ factors, as shown here, the relevant structural module is the σR2/4 linker (Figs. 3 and 4; Supplementary Fig. 3); for group-1, group-2, and group-3 σ factors, the module is the functionally analogous—but not structurally homologous—σR3/4 linker2,3,5,6,12–19,22, and for group-544/σ^II^ σ factors, the module is the functionally analogous—but not structurally homologous—region II.3 (RII.3)20,21.

More broadly, our results, in conjunction with previous results, indicate that cellular transcription initiation complexes in all organisms—bacteria, archaea, and eukaryotes—contain structural modules that enter the RNAP active-center cleft to interact with template-strand ssDNA and then leave the RNAP active-center cleft by threading through the RNAP RNA-exit channel. In different classes of bacterial transcription initiation complexes, as described in the preceding paragraph, these roles are performed by the functionally analogous—but not structurally homologous—σR2/4 linker, σR3/4 linker, and RII.3.12,5,6,7;12,19–22. In archaeal transcription initiation complexes, these roles are performed by the TFB zinc ribbon and CSB, which are unrelated to the σR2/4 linker, σR3/4 linker, and RII.3.37. In eukaryotic RNAP-I-, RNAP-II-, and RNAP-III-dependent transcription initiation complexes, these roles are performed by the Rrn7 zinc ribbon and B-reader, the TFIIB zinc ribbon and B-reader, and the Brf1 zinc ribbon,
respectively, each of which is unrelated to the σR2/4 linker, σR3/4 linker, and RLH. It is extraordinary that non-homologous, structurally and phylogenetically unrelated, structural modules are used to perform the same roles in different transcription initiation complexes, and is unknown how or why this occurs.

Our results define the protein–DNA interactions that ECF σ factor σε uses to recognize the -10 element of a σε-dependent promoter. The consensus sequence obtained in this work for the 10-element of a σε-dependent promoter, T−12−G−11−N−10−C/A−9−G−10−C/A−2−G−7 (Fig. 6b), confirms and extends the literature-consensus sequence35,36, and the structural data of this work account for specificity at each specified position of the consensus sequence (Figs. 3–5; Supplementary Fig. 4).

Previous work indicates that RNAP-σL holoenzyme prefers a C-G sequence immediately upstream of the -10-element (C-G-A)28,59, indicating that this preference may be shared by many RNAP-L-factor-dependent transcription initiation complexes containing chimeric factors57,60 comprising or2 of a Mtb σ factor of interest fused to the σR2/4 linker through or4 of Mtb σl (Supplementary Fig. 7B; left red arrow) and containing the promoter sequence for the Mtb σ factor of interest. In the crystal form identified and analyzed in this work, or2 of each molecule of transcription initiation complex makes no interactions with other molecules of transcription initiation complex in the crystal lattice (Supplementary Fig. 7A), and, therefore, with this crystal form, it may be possible to substitute or2 without losing the ability to form crystals (Supplementary Fig. 7A). This potentially provides a platform for systematic structural analysis of or2 and or2-DNA interactions for the 13 Mtb σ factors, by determination of crystal structures of transcription initiation complexes containing chimeric factors57,60, comprising or2 of a Mtb σ factor of interest fused to the σR2/4 linker through or4 of Mtb σl (Supplementary Fig. 7B; left red arrow) and containing the promoter sequence for the Mtb σ factor of interest.

In the crystal form identified and analyzed in this work, there also are no lattice interactions for the connector between or2 and or4, and it appears likely there would be no lattice interactions even if that connector were to contain or3 and a or3/4 linker, as in group-1, group-2, and group-3 σ factors (Supplementary Figure 7A). Accordingly, this crystal form potentially provides a platform for systematic structural analysis not only of or2 and its protein-DNA interactions, but also of the connector between or2 and or4 and its protein-DNA interactions, for the 13 Mtb σ factors, by determination of crystal structures of transcription initiation complexes containing chimeric σ factors comprising or2 and the connector of one Mtb σ factor fused to or4 of Mtb σl (Supplementary Figure 7B, right red arrow) and containing the promoter sequence for the Mtb σ factor of interest.

Methods

M. tuberculosis RNAP core enzyme. Mtb RNAP core enzyme was prepared by co-expression of genes for Mtb RNAP β subunit, RNAP β subunit, N-terminally decapeptide-tagged RNAP σ subunit, and RNAP ω subunit in E. coli (plasmids pACYC-ρOa, pCOLADuet-ρOa, pCDE-ρOc, and pCDE-ρOa)61 (gift of J. Mukhopadhyay, Bose Institute, Kolkata, India); E. coli strain BL21(DE3) (Invitrogen), followed by cell lysis, polyethyleneimine precipitation, ammonium sulfate precipitation, immobilized-metal-ion affinity chromatography on Ni-NTA agarose (Qiagen), and anion-exchange chromatography on Mono Q (GE Healthcare), as described19.

M. tuberculosis RNAP ω. Mtb RNAP ω was prepared by expression of a gene for N-terminally hexahistidine-tagged Mtb σω in E. coli (plasmid pET30a-Mtb-σω62 (gift of S. Rodrigue, Université de Sherbrooke, Canada); E. coli strain BL21(DE3) (Invitrogen), followed by cell lysis, immobilized-metal-ion affinity chromatography on Ni-NTA agarose (Qiagen), and anion-exchange chromatography on Mono Q (GE Healthcare), as described19.

M. tuberculosis σε. Mtb RNAP σε was prepared by expression of a gene for N-terminally hexahistidine-tagged Mtb σε in E. coli, followed by cell lysis, immobilized-metal-ion affinity chromatography on Ni-NTA agarose (Qiagen), and anion-exchange chromatography on Mono Q (GE Healthcare), as described19.
centrifuged (20,000×g, 30 min at 4 °C), the pellet was re-suspended in buffer B (8 M urea, 10 mM Tris–HCl, pH 7.9, 10 mM MgCl₂, 1 mM ZnCl₂, 1 mM EDTA, 10 mM 2-mercaptoethanol, and the previous), and the centrifuged, and eluted with 50 µl of buffer B containing 200 mM imidazole. The sample was subjected to step dialysis for renaturation [10 kDa MWCO Amicon Ultra-15 centrifugal ultrafilters (EMD Millipore); dialysis 4 h at 4 °C against 8 volumes 5 mM MgCl₂, and 1 mM 2-mercaptoethanol for 12 h at 4 °C. The reaction mixture containing 12

5.5, 200 mM sodium acetate, and 10% PEG4000; 22 °C) yielded high-quality, rod-like crystals with dimensions of 0.4 mm × 0.1 mm × 0.1 mm in 2 weeks (Supplementary Figs. 1–2). Crystals were transferred to reservoir solution containing 18% (v/v) (2 R-3 i-(−)-2,3-butanediol (Sigma-Aldrich) and flash-cooled with liquid nitrogen. Analogous procedures were used for Mtb RPItc5-δ₅₋₆sp4, RPItc5-δ₅₋₆sp5, RPItc5-δ₅₋₆sp6, [BrU]RPO-δ₅₋₆sp6, and [MeMet17,58] RPO-δ₅₋₆sp6.

Diffraction data collection. Diffraction data were collected from cryo-cooled crystals at Argonne National Laboratory beamline 19ID-D and Stanford Synchrotron Radiation Lightsource SSSL-9.2. Data were processed using HKL200064. The resolution cut-off criteria were: (i) f(0) = 1.0, (ii) CC anh (highest resolution shell) >0.5.

Structure determination and structure refinement. The structure of Mtb RPItc5-δ₅₋₆sp was solved by molecular replacement with MOLREP65 using the structure of Mtb RPO (PDB SUHA)66, omitting δ₅ and nucleic acids, as the search model. One molecule of RNAP was in present in the asymmetric unit. Early-stage refinement included rigid-body refinement of RNAP core enzyme, followed by rigid-body refinement of each subunit of RNAP core enzyme, followed by rigid-body refinement of 38 domains of RNAP core enzyme (methods as described67). Electron density for δ₅ and nucleic acids was unambiguous, but was not included in models in early-stage refinement. Cycles of iterative model building with Coot68 and refinement with Phenix67 then were performed. Improvement of the coordinate maps was achieved by density shading, and electron density for δ₅ and nucleic acids, which were not included in models at this stage, improved over successive cycles. δ₅ and nucleic acids then were built into the model and refined in a stepwise fashion. The final model was generated by Xyz-coord-re refinement with secondary-structure restraints, followed by group B-factor and individual B-factor refinement. The final model, refined to ρM (ρM of 0.19 and 0.23, respectively, was deposited in the PDB with accession code 6DVC (Table 1).

Analogous procedures were used to solve and preliminarily refine structures of Mtb RPItc5-δ₅₋₆sp, RPItc5-δ₁₋₆sp, RPItc5-δ₅₋₆sp6, and [BrU]RPO-δ₅₋₆sp6; models of δ₅ and nucleic acids then were built into m2, D2F, difference maps, and additional cycles of refinement and model building were performed. The final models were deposited in the PDB with accession codes 6DV9, 6DVB, and 6DVE (Table 1).

Distance cut-offs for assignment of H-bonds and van der Waals interactions were 3.5 and 4.5 Å, respectively.

Transcription assays. For transcription experiments in Fig. 6 and Supplementary Figs. 1E and 6, reaction mixtures contained (10 µl): 75 nM Mtb RNAP δ₅₋₆holoenzyme or Mtb RNAP δ₅₋₆holoenzyme derivative, 25 nM [α⁻³²P]UTP (0.003 Bq/fmol), 100 µM ATP, and 100 µM GTP in transcription buffer (40 µM Tris–HCl, pH 8.0, 75 mM NaCl, 5 mM MgCl₂, 2.5 mM EDTA, and 12.5% glycerol). Reaction components other than DNA and nucleotides were pre-incubated 5 min at 22 °C. DNA was added and reaction mixtures were incubated 5 min at 37 °C; nucleotides were added and reaction mixtures were further incubated 5 min at 37 °C. Products were terminated by addition of 2× SDS-PAGE loading buffer (0.8 M urea, 10 mM EDTA, 0.04% bromophenol blue, and 0.04% xylene cyanol). Products were heated 5 min at 95 °C, cooled 5 min on ice, and applied to 16% polyacrylamide (19:1 acrylamide/bisacrylamide, 7 M urea) slab gels (Bio-Rad), electrophoresed in TBE (90 mM Tris–borate, pH 8.0, and 0.2 mM EDTA), and analyzed by storage-phosphor scanning (Typhoon; GE Healthcare). Relative transcriptional activities were calculated from yields of full-length RNA products.

Transcription experiments in Supplementary Fig. 1F were performed in the same manner as transcription experiments in Fig. 6 and Supplementary Figs. 1E and 6, but using reaction mixtures containing (10 µl): 600 nM Mtb RNAP δ₅₋₆holoenzyme, 400 nM annealed nontemplate and template strands of nucleic-acid fragments of δ₅₋₆sp scaffold, 100 µM ATP, 100 µM GTP, and 40 µM Tris–HCl, pH 8.0, 75 mM NaCl, 5 mM MgCl₂, 2.5 mM EDTA, and 12.5% glycerol. Reaction components other than DNA and nucleotides were pre-incubated 5 min at 22 °C. DNA was added and reaction mixtures were incubated 5 min at 37 °C; nucleotides were added and reaction mixtures were further incubated 5 min at 37 °C. Products were terminated by addition of 2× SDS-PAGE loading buffer (0.8 M urea, 10 mM EDTA, 0.04% bromophenol blue, and 0.04% xylene cyanol). Products were heated 5 min at 95 °C, cooled 5 min on ice, and applied to 16% polyacrylamide (19:1 acrylamide/bisacrylamide, 7 M urea) slab gels (Bio-Rad), electrophoresed in TBE (90 mM Tris–borate, pH 8.0, and 0.2 mM EDTA), and analyzed by storage-phosphor scanning (Typhoon; GE Healthcare). Relative transcriptional activities were calculated from yields of full-length RNA products.

Transcription experiments in Supplementary Fig. 1E were performed in the same manner as transcription experiments in Fig. 6 and Supplementary Figs. 1E and 6.
References

1. 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).

2. Murakami, K., Masuda, S. & Darst, S. Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. Science 296, 1280–1284 (2002).

3. Vassylyev, D. et al. Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. Nature 417, 712–719 (2002).

4. Meeker, V. et al. Structural organization of bacterial RNA polymerase holoenzyme and the RNA polymerase-promoter open complex. Cell 108, 599–612 (2002).

5. Kulbachinskaya, 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).

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

7. Basu, R. et al. Structural basis of transcription initiation by bacterial RNA polymerase holoenzyme. J. Biol. Chem. 289, 24549–24559 (2014).

8. Pupov, D., Kuzin, L., 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).

9. Duchi, D. et al. RNA polymerase pausing during initial transcription. Mol. Cell 63, 939–950 (2016).

10. Lerner, E. et al. Backtracked and paused transcription initiation intermediate of Escherichia coli RNA polymerase. Proc. Natl Acad. Sci. USA 113, E6556–E6571 (2016).

11. Dulin, D. et al. Pausing controls branching between productive and non-productive pathways during initial transcription in bacteria. Nat. Commun. 9, 1478 (2018).

12. Murakami, K. The x-ray crystal structure of Escherichia coli RNA polymerase sigma70 holoenzyme. J. Biol. Chem. 288, 9126–9134 (2013).

13. Bae, B. et al. Phage T7 gp2 inhibitor of Escherichia coli RNA polymerase involves misappropriation of sigma70 domain 1.1. Proc. Natl Acad. Sci. USA 110, 19772–19777 (2013).

14. Zhang, Y. et al. GE23077 binds to the RNA polymerase ρ- and ρ+ sites and prevents the binding of initiating nucleotides. eLife 3, e02450 (2014).

15. Zuo, Y. & Steitz, T. Crystal structures of the E. coli transcription initiation complexes with a complete bubble. Mol. Cell 58, 534–540 (2015).

16. Bae, B., Feklistov, A., Lass-Napiorkowska, A., Landick, R. & Darst, S. Structure of a bacterial RNA polymerase holoenzyme open promoter complex. eLife 4, e08504 (2015).

17. Feng, Y., Zhang, Y. & Ebright, R. H. Structural basis of transcription pausing during transcription. Science 352, 1330–1333 (2016).

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

19. Lin, W. et al. Structural basis of Mycobacterium tuberculosis transcription and transcription inhibition. Mol. Cell 66, 169–179 (2017).

20. Yang, Y. et al. Structures of the RNA polymerase sigma34 reveal new and conserved regulatory strategies. Science 349, 882–885 (2015).

21. Glyde, R. et al. Structures of RNA polymerase closed and intermediate complexes reveal mechanisms of DNA opening and transcription initiation. Mol. Cell 67, 110–116 (2017).

22. Liu, B., Zuo, Y. & Steitz, T. A. Structures of E. coli sigma5 transcription initiation complexes provide new insights into polymerase mechanism. Proc. Natl Acad. Sci. USA 113, 4051–4056 (2016).

23. Lonetto, M., Brown, K., Rudd, K. E. & Buttnner, M. Analysis of the Streptomyces coelicolor sigE gene reveals the existence of a subfamily of eubacterial RNA polymerase sigma factors involved in the regulation of extrapolytopic functions. Proc. Natl Acad. Sci. USA 91, 7573–7577 (1994).

24. Missiakas, D. & Raina, S. The extrapolytopic function sigma factors: role and regulation. Mol. Microbiol. 28, 1059–1066 (1998).

25. Helmann, J. The extrapolytopic function (ECF) sigma factors. Adv. Microb. Physiol. 46, 47–110 (2002).

26. Staron, A. et al. The third pillar of bacterial signal transduction: classification of the extrapolytopic function (ECF) sigma factor protein family. Mol. Microbiol. 74, 557–581 (2009).

27. Mascher, T. Signaling diversity and evolution of extrapolytopic function (ECF) sigma factors. Curr. Opin. Microbiol. 16, 148–155 (2013).

28. Campagne, S., Allain, F. & Vorholt, J. Extra cytoplasmic sigma factor functions, recent structural insights into promoter recognition and regulation. Curr. Opin. Struct. Biol. 30, 71–78 (2015).

29. Helmann, J. Bacillus subtilis extrapolytopic function (ECF) sigma factors and defense of the cell envelope. Curr. Opin. Microbiol. 30, 122–132 (2015).

30. Sineva, E., Savkina, M. & Ades, S. Themes and variations in gene regulation by extrapolytopic function (ECF) sigma factors. Curr. Opin. Microbiol. 36, 128–137 (2017).

31. Manganelli, R. et al. Sigma factors and global gene regulation in Mycobacterium tuberculosis. J. Bacteriol. 186, 895–902 (2004).

32. Rodrigue, S., Provvedi, R., Jacques, P., Gaudreau, L. & Manganelli, R. The sigma factors of Mycobacterium tuberculosis. FEBS Microbiol. Rev. 30, 926–941 (2006).

33. Sachdeva, P., Misra, R., Tyagi, A. & Singh, V. The sigma factors of Mycobacterium tuberculosis: regulation of the regulators. FEBS J. 277, 605–626 (2010).

34. Newton-Foot, M. & Gey van Pittius, N. C. The complex architecture of mycobacterial promoters. Tuberculosis 93, 60–74 (2013).

35. Manganelli, R. Sigma factors: key molecules in Mycobacterium tuberculosis physiology and virulence. Microbiol. Spectr. 2, MGC2-0007-2013 (2014).

36. Hahn, M., Ramani, S., Anaya, M. & Hasson, P. The Mycobacterium tuberculosis extrapolytopic function-sigma factor Sigl regulates polyketide synthases and secreted or membrane proteins and is required for virulence. J. Bacteriol. 187, 7062–7071 (2005).

37. Dainese, E. et al. Post translational regulation of Mycobacterium tuberculosis extrapolytopic function sigma factor sigma S and roles in virulence and in global regulation of gene expression. Infect. Immun. 74, 2457–2461 (2006).
38. Rodrígue, S. et al. Identification of mycobacterial sigma factor binding sites by chromatin immunoprecipitation assays. J. Bacteriol. 189, 1505–1513 (2007).

39. Feklistov, A. & Darst, S. Structural basis for promoter-10 element recognition by the bacterial RNA polymerase sigma subunit. Cell 147, 1257–1269 (2011).

40. Lim, H., Lee, H., 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).

41. Cowling, J.H. High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. Science 244, 1081–1085 (1989).

42. Ebright, R. Use of loss-of-contact substitutions to identify residues involved in an amino acid–base pair contact: effect of substitution of Gln18 of Lac repressor by Gly, Ser, and Leu. J. Biomol. Struct. Dyn. 3, 281–297 (1985).

43. Elsholtz, M. Evidence for a contact between glutamine-18 of lac repressor and base pair 7 of lac operator. Proc. Natl Acad. Sci. USA 83, 303–307 (1986).

44. Ebright, R. Identification of amino acid–base pair contacts by genetic methods. Methods Enzymol. 208, 620–640 (1991).

45. Zhang, X. & Ebright, R. Identification of a contact between arginine-180 of the catabolite gene activator protein (CAP) and base pair 5 of the DNA site in the CAP-DNA complex. Proc. Natl Acad. Sci. USA 87, 4717–4721 (1990).

46. Vahedian-Movahed, H., Ebright, R. & forcefully, R. Use of loss-of-contact substitutions to identify residues involved in an amino acid–base pair contact: effect of substitution of Gln18 of Lac repressor by Gly, Ser, and Leu. J. Biomol. Struct. Dyn. 3, 281–297 (1985).

57. Rhodius, V. et al. Design of orthogonal genetic switches based on a crosstalk map on gel. EMBO J. 26, 269–2709 (2007).

58. Kumar, A. et al. A hybrid sigma subunit directs RNA polymerase to a hybrid promoter in Escherichia coli. J. Mol. Biol. 246, 583–571 (1995).

59. Campagne, S., Marsh, M., Capitan, G., Vorholt, J. A. & Allain, F. Structural basis for -10 promoter element melting by environmentally induced sigma factors. Nat. Struct. Mol. Biol. 21, 269–276 (2014).

60. Koo, B., Rhodius, V., Campbell, E. & Gross, C. Mutational analysis of the lac promoter using a hybrid sigma subunit. EMBO J. 26, 269–2709 (2007).

61. Banerjee, R., Rudra, P., Prajapati, R., Sengupta, S. & Mukhopadhyay, J. Optimization of recombinant Mycobacterium tuberculosis RNA polymerase expression and purification. Tuberculosis 94, 397–404 (2014).

62. Koo, B., Rhodius, V., Campbell, E. & Gross, C. Mutational analysis of the lac promoter using a hybrid sigma subunit. EMBO J. 26, 269–2709 (2007).

63. Stols, L., Millard, C., Dementiev, I. & Donnelly, M. Production of selenomethionine-labeled proteins in two-liter plastic bottles for structure determination. J. Struct. Funct. Genomics 5, 95–102 (2004).

64. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. 276, 307–326 (1997).

65. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D 54, 760–763 (1998).

66. Emley, P., Lohkamp, B., Scott, W. & Cowtan, K. Features and development of Coot. Acta Crystallogr. D 66, 486 (2010).

67. Adams, P. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D 66, 213 (2010).

68. Kamerer, W., Deuschle, U., Gentz, R. & Bujard, H. Functional dissection of Escherichia coli promoters: information in the transcribed region is involved in late steps of the overall process. EMBO J. 8, 2995–3000 (1986).

69. Dickson, R., Abelson, J., Barnes, W. & Reznikoff, W. Genetic regulation: the lac control region. Science 187, 17–35 (1975).

70. Tarnell, W. & Roberts, J. Mechanism of intrinsic transcription termination and antitermination. Science 284, 611–615 (1999).

71. Workman, C. et al. enOGOS: a versatile web tool for energy normalized sequence logos. Nucleic Acids Res. 33, W389–W392 (2005).

Acknowledgements
This work was supported by NIH Grant GM041376 to R.H.E. The authors thank J. Mukhopadhyay and S. Rodrigue for plasmids and APS at Argonne National Laboratory and Stanford Synchrotron Radiation Lightsource for beamline access.

Author contributions
W.L. and M.S.C. prepared RNAP derivatives. W.L., Y.F. and K.D. performed structure determination. W.L., D.D. and S.M. performed sequence analyses and biochemical experiments. R.H.E. designed the study, analyzed data, and wrote the paper.

Additional information
Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-019-08443-3.

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 the anonymous reviewers 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