THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 288, NO. 13, pp. 9126–9134, March 29, 2013

X-ray Crystal Structure of *Escherichia coli* RNA Polymerase σ^{70} Holoenzyme*[S]

Received for publication, November 7, 2012, and in revised form, January 25, 2013 Published, JBC Papers in Press, February 6, 2013, DOI 10.1074/jbc.M112.430900

Katsuhiko S. Murakami[1]

From the Department of Biochemistry and Molecular Biology, Center for RNA Molecular Biology, Pennsylvania State University, University Park, Pennsylvania 16802

Background: A crystal structure of *Escherichia coli* RNA polymerase (RNAP) has not been determined.
Results: The $\sigma_{1.1}$ and α subunit C-terminal domain structures have been determined in the context of an intact RNAP.
Conclusion: $\sigma_{1.1}$ localizes within the RNAP DNA-binding channel and must disengage from this site to form an open complex.
Significance: This work enables future structure determination of bacterial RNAP mutants.

***Escherichia coli* RNA polymerase (RNAP) is the most studied bacterial RNAP and has been used as the model RNAP for screening and evaluating potential RNAP-targeting antibiotics. However, the x-ray crystal structure of *E. coli* RNAP has been limited to individual domains. Here, I report the x-ray structure of the *E. coli* RNAP σ^{70} holoenzyme, which shows σ region 1.1 ($\sigma_{1.1}$) and the α subunit C-terminal domain for the first time in the context of an intact RNAP. $\sigma_{1.1}$ is positioned at the RNAP DNA-binding channel and completely blocks DNA entry to the RNAP active site. The structure reveals that $\sigma_{1.1}$ contains a basic patch on its surface, which may play an important role in DNA interaction to facilitate open promoter complex formation. The α subunit C-terminal domain is positioned next to σ domain 4 with a fully stretched linker between the N- and C-terminal domains. *E. coli* RNAP crystals can be prepared from a convenient overexpression system, allowing further structural studies of bacterial RNAP mutants, including functionally deficient and antibiotic-resistant RNAPs.**

RNA polymerase (RNAP)[2] is the central enzyme of gene expression, and all life forms have RNAPs that function as multisubunit protein complexes (multisubunit cellular RNAP). The common core of the multisubunit RNAPs is composed of five subunits that are conserved in bacteria, archaea, and eukaryotes. Bacterial RNAP is the simplest form of this family (composed of the core enzyme $\alpha_I\alpha_{II}\beta\beta'\omega$ subunits), whereas in eukaryotes and archaea, RNAP possesses additional polypeptides to form 11∼15-subunit complexes (1).

In bacteria, one of several different σ factors binds to the core enzyme to form the holoenzyme, which is responsible for recognizing promoter DNA. σ^{70} in *Escherichia coli* and SigA in other bacteria belong to the group 1 (primary or housekeeping) σ factor family (2). These σ factors contain distinct regions of highly conserved amino acid sequence and are composed of four domains: $\sigma_{1.1}$ (region 1.1), σ2 (regions 1.2–2.4), σ3 (regions 3.0–3.2), and σ4 (regions 4.1–4.2) (3). Group 1 σ factors can bind to promoter DNA as part of the holoenzyme; once it binds to the core enzyme, the σ2, σ3, and σ4 domains are ideally positioned to recognize the promoter DNA sequences of −10, extended −10, and −35, respectively (4, 5).

In addition to the σ2, σ3, and σ4 domains, the group 1 σ family contains an ∼100-amino acid N-terminal extension, $\sigma_{1.1}$, which is a negatively charged α helical domain (6). The $\sigma_{1.1}$ domain has been shown to accelerate the formation of the open complex at some promoters and suggested to reside inside the RNAP main channel (7). This channel is positively charged to accommodate nucleic acids in the open complex and the transcription elongation complex. It has been proposed that during open complex formation, signals from DNA may induce opening and closing of the RNAP clamp, causing $\sigma_{1.1}$ to eject from the RNAP main channel (4, 8). Given its flexible nature, $\sigma_{1.1}$ has not been solved in all *Thermus* RNAP holoenzyme crystal structures that have been reported (5, 9–12). Only an NMR structure of $\sigma_{1.1}$ from *Thermotoga maritima* has been reported, and it consists of three α helices with a compact hydrophobic core formed by highly conserved hydrophobic residues (6).

Since the first discovery of RNAP in the early 1960s (13), the RNAP from *E. coli* has been the primary model system of choice for understanding functions of cellular RNAPs for many reasons. For example, active *E. coli* RNAP can be conveniently reconstituted *in vitro* from its individual subunits using either wild-type or mutant proteins (14, 15), and its mechanism can be easily probed *in vitro* in the presence of purified template DNA, σ factors, and transcription factors. A simple and robust *E. coli* transcription system also makes it an excellent model for single-molecule studies of RNAPs (16).

X-ray crystal structures of bacterial RNAPs have been determined only from the *Thermus* genus. Because of the high sequence conservation among RNAPs from all species of bacteria, the most insight derived from the *Thermus* RNAP has been generalized to represent the transcription apparatus in all bacteria (4, 5, 9–12, 17–19). Nevertheless, without the structure of *E. coli*

* This work was supported, in whole or in part, by National Institutes of Health Grant GM087350-A1. This work was also supported by a research fellowship from the Center of Excellence of the Japan Society for the Promotion of Science and a long-term postdoctoral research fellowship from the Human Frontier Science Program.
[S] This article contains supplemental Movies S1–S4.
The atomic coordinates and structure factors (code 4IGC) have been deposited in the Protein Data Bank (http://wwpdb.org/).
[1] To whom correspondence should be addressed. Tel.: 814-865-2758; E-mail: kum14@psu.edu.
[2] The abbreviations used are: RNAP, RNA polymerase; βi, β insert; *Eco*β, *E. coli* β subunit; *Taq*β, *T. aquaticus* β subunit; TLH, trigger loop/helix; α-CTD, α subunit C-terminal domain; α-NTD, α subunit N-terminal domain.

RNAP available, it is difficult to fully interpret the enormous amount of data that have been collected on *E. coli* RNAP. The structure of *E. coli* RNAP will also generate new insight about structural domains and motifs, as well as interactions with some ligands (*e.g.* ppGpp) and antibiotics (*e.g.* lipiarmycin) that specifically affect *E. coli* but not the *Thermus* RNAPs (20, 21). These structural insights are important to identify their binding sites and to understand the mechanisms of action.

EXPERIMENTAL PROCEDURES

Preparation and Crystallization of the E. coli RNAP Holoenzyme—The polycistronic plasmid pGEMABC was created for overexpressing the *rpoA* (encoding the α subunit), *rpoB* (encoding the β subunit), and *rpoC* (encoding the β′ subunit) genes as follows. The plasmid pGEMA185 expressing *rpoA* under the control of an IPTG-inducible T7 RNAP promoter (22) was digested at a BamHI site located downstream of *rpoA*. A DNA fragment containing the *rpoB-rpoC* genes was isolated from the pPNE2017 plasmid[3] by BamHI digestion and inserted at the BamHI site of pGEMA185. pGEMABC expresses a single mRNA containing the *rpoA-rpoB-rpoC* genes.

All core RNAP subunits were expressed in *E. coli* BL21(DE3) cells transformed with pGEMABC (encoding *rpoA*, *rpoB*, and *rpoC*) and pACYCDuet-1_Ec_rpoZ (encoding *rpoZ*). Core RNAP was purified as follows. ~16 g of cell paste was suspended in 50 ml of lysis buffer (50 mM Tris-HCl (pH 8 at 4 °C), 1 mM EDTA, 5 mM 2-mercaptoethanol, 1× protease inhibitor mixture, and 2 mM PMSF), and cells were lysed using an EmulsiFlex C3 homogenizer (Avestin Inc.) at 20,000 p.s.i. After a low-speed spin, RNAP in the soluble fraction was precipitated by adding 10% polyethyleneimine (Polymin P) solution (final concentration of 0.6%), and the pellet was recovered by low-speed centrifugation. RNAP was eluted from the pellet by suspension in TGED buffer (10 mM Tris-HCl (pH 8 at 4 °C), 10% glycerol, 0.1 mM EDTA, and 2 mM DTT) + 1 M NaCl and then precipitated by ammonium sulfate (final 60% saturation). The pellet was suspended in TGED buffer and dialyzed against TGED buffer + 50 mM NaCl. Core RNAP was purified by Bio-Rex 70 (Bio-Rad), Resource Q (GE Healthcare), and Superdex 200 (GE Healthcare) column chromatography. *E. coli* σ^{70} was expressed in BL21(DE3) cells transformed with pGEMD (22). After cells were lysed by sonication, σ^{70} was purified by HiTrap Q HP (GE Healthcare) and Superdex 200 column chromatography.

The RNAP holoenzyme was prepared by adding a 3-fold excess of σ^{70} to core RNAP, followed by incubation at 30 °C for 30 min and purification by Superdex 200 column chromatography. Crystals were obtained by hanging drop vapor diffusion by mixing equal volumes of RNAP holoenzyme solution (~20 mg/ml) and crystallization solution (0.1 M HEPES-HCl (pH 7.0), 0.2 M calcium acetate, and ~15% PEG 400) and incubating at 22 °C over the same crystallization solution. For cryocrystallography, crystals were soaked in crystallization solution containing 25% PEG 400. Selenomethionyl-substituted proteins, including core RNAP and σ^{70}, were prepared by suppression of methionine biosynthesis (23). The crystals belong to the primitive orthorhombic space group (Table 1) containing two 440-kDa RNAP holoenzymes per asymmetric unit, and these RNAPs have almost identical structures (0.643-Å root mean square deviations by a structure alignment using the β′ subunit), with some minor deviations in the position of the σ nonconserved and σ4 domains.

TABLE 1
Data collection and refinement statistic of the *E. coli* RNA polymerase σ^{70} holoenzyme
Data sets were collected at MacCHESS beamline A1. PDB, Protein Data Bank; r.m.s.d., root mean square deviations.

PDB code	4IGC
Data collection	
Space group	$P2_12_12_1$
Cell dimensions (Å)	
a	187.308
b	205.901
c	309.185
Resolution (Å)	30–3.60
Total reflections	592,860
Unique reflections	123,448
Redundancy	4.5 (2.9)[a]
Completeness (%)	96.81 (93.67)[a]
I/σ	10.13 (2.04)[a]
R_{sym}	0.091 (0.469)[a]
Refinement	
Resolution (Å)	30–3.70
R_{work}	0.242
R_{free}	0.285
r.m.s.d.	
Bond length (Å)	0.003
Bond angles	0.77°

[a] The highest resolution shell (3.66 to 3.6 Å) is shown in parentheses.

X-ray Data Collections and Structure Determination—The native data set was collected at Macromolecular Diffraction at the Cornell High Energy Synchrotron Source (MacCHESS) beamline A1 (Cornell University, Ithaca, NY). The data sets of SeMet-labeled crystals were collected at Berkley Center for Structural Biology (BCSB) beamline 8.2.1 (Lawrence Berkeley National Laboratory, Berkley, CA). The *E. coli* core RNAP model (24) was used as a search model for the molecular replacement (25). The data were processed by HKL2000 (26). Anomalous signals from SeMet were located by phase obtained from molecular replacement. Rigid body refinements were performed, and further adjustments to the model were performed manually. The resulting model phases allowed me to position *E. coli* σ^{70} structures (27, 28) in the electron density map. Positional refinement with non-crystallographic symmetry and secondary structure restraints was performed using the program PHENIX (29), and deformable elastic network (DEN) refinement was performed using Crystallography & NMR System (CNS) version 1.3 (30). The resulting map allowed segments that were not present in the search model to be built manually by Coot (31). The final coordinates and structure factors were submitted to the Protein Data Bank with code 4IGC.

RESULTS AND DISCUSSION

E. coli σ^{70} RNAP Holoenzyme Preparation and Crystallization—Endogenous *E. coli* RNAP can be purified from cells by a combination of RNAP-DNA co-precipitation using Polymin P and column chromatography (32). However, the yield and purity of endogenous *E. coli* RNAP are inadequate to

[3] N. Fujita and R. E. Glass, personal communication.

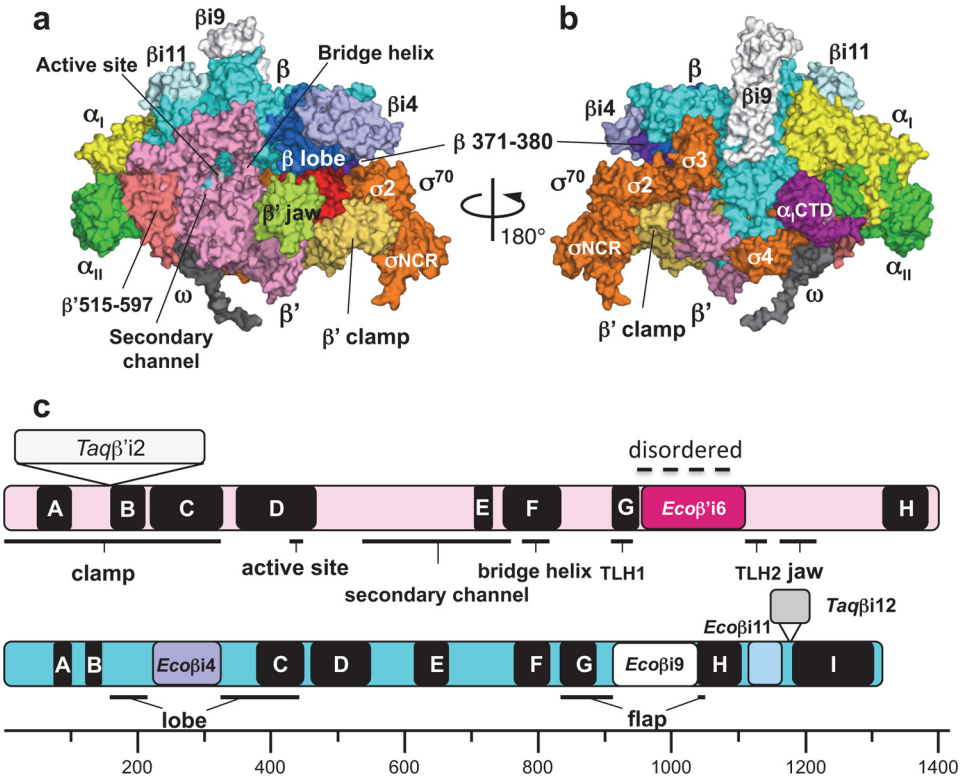

FIGURE 1. **Three-dimensional crystal structure of the *E. coli* RNAP σ^{70} holoenzyme.** *a* and *b*, surface representation of the *E. coli* RNAP holoenzyme. *Panel a* shows a view from the RNAP secondary channel leading to the active site, and *panel b* shows the σ-binding site. Each subunit of RNAP is denoted by a unique color: *yellow*, α_I; *green*, α_{II}; *cyan*, β; *pink*, β'; *gray*, ω; and *orange*, σ^{70}. Several domains described under "Results and Discussion" are also denoted by a unique color and are indicated. *c*, linear maps of the β' (*upper*) and β (*lower*) subunits. Conserved regions of the β' subunit (A–H) and the β subunit (A–I) are shown as *black boxes* with the structural domains of RNAP. Specific insertions of the *E. coli* and *Thermus* RNAPs are shown by the same colors as in *panels a* and *b* and in Fig. 2.

obtain high-quality crystals for x-ray crystallography (<1 mg of RNAP is generated from 1 liter of cell culture). Therefore, I developed a co-overexpression plasmid (pGEMABC) that expresses the *rpoA* (encoding the α subunit), *rpoB* (encoding the β subunit), and *rpoC* (encoding the β' subunit) genes under a single T7 RNAP promoter. This overexpression system drastically improves the yield and purity of RNAP (10 mg of RNAP from 1 liter of cell culture). The σ^{70} holoenzyme can be prepared by adding recombinant σ^{70} to core RNAP. Both the core and holoenzyme formed crystals, but neither diffracted beyond 10 Å resolution. pGEMABC overexpresses the α, β, and β' subunits but not the ω subunit; thus, purified RNAP contains a substoichiometric amount of the ω subunit. The importance of the ω subunit for RNAP assembly and formation was suggested by a biochemical experiment (33) and by the *Thermus* RNAP crystal structure, which shows that the ω subunit binds the C-terminal tail of the β' subunit (see Fig. 3*b* and supplemental Movie S4) (17).

To prepare RNAP containing a stoichiometric amount of the ω subunit, all RNAP subunits were overexpressed by pGEMABC and pACYCDuet-1_Ec_rpoZ, which overexpresses the ω subunit. The *E. coli* RNAP holoenzyme was prepared *in vitro* by addition of σ^{70}, which produced better quality crystals that allowed determination of the structure by x-ray crystallography.

Structure Determination of the E. coli σ^{70} RNAP Holoenzyme—The crystals contain two 440-kDa RNAP holoenzyme molecules, designated $RNAP_A$ and $RNAP_B$, per asymmetric unit. The structure was solved by molecular replacement with an *E. coli* RNAP core enzyme model (24). After density modification, the resulting electron density map had several deviations from the molecular replacement solution, including the following regions: 1) β insert 4 (βi4, residues 225–343, previously named β dispensable region 1/βDR1/SI1), 2) β insert 9 (βi9, residues 938–1042, previously named β dispensable region 2/βDR2/SI2), 3) β insert 11 (βi11, residues 1122–1180, present between β conserved regions H and I), 4) β' insert 6 (β'i6, residues 942–1129, present in the middle of the highly conserved β' trigger loop/helix), 5) β' residues 515–597 (present between β' conserved regions B and C), and 6) the C-terminal tails of the β' and ω subunits (Fig. 1; see Fig. 3*a*). The overall structures of βi4 and βi9 are similar to the structures in the previously reported *E. coli* RNA core enzyme model (24), but their orientations relative to the main body of the RNAP are different. The crystal structures of the *E. coli* σ^{70} domains (27, 28) were manually placed in the $F_o - F_c$ map, resulting in good fits of σ2, σ3, and σ4. Anomalous signals from SeMet sites from both the core enzyme and σ^{70} were used as guides for model building and refinement.

Structure of the E. coli σ^{70} RNAP Holoenzyme—The overall structure of the *E. coli* RNAP holoenzyme is similar to the structure of *Thermus* RNAP, resembling a crab claw with two pincers that constitute the DNA-binding cleft and the active site (Fig. 1 and supplemental Movie S1) (17). The β' subunit

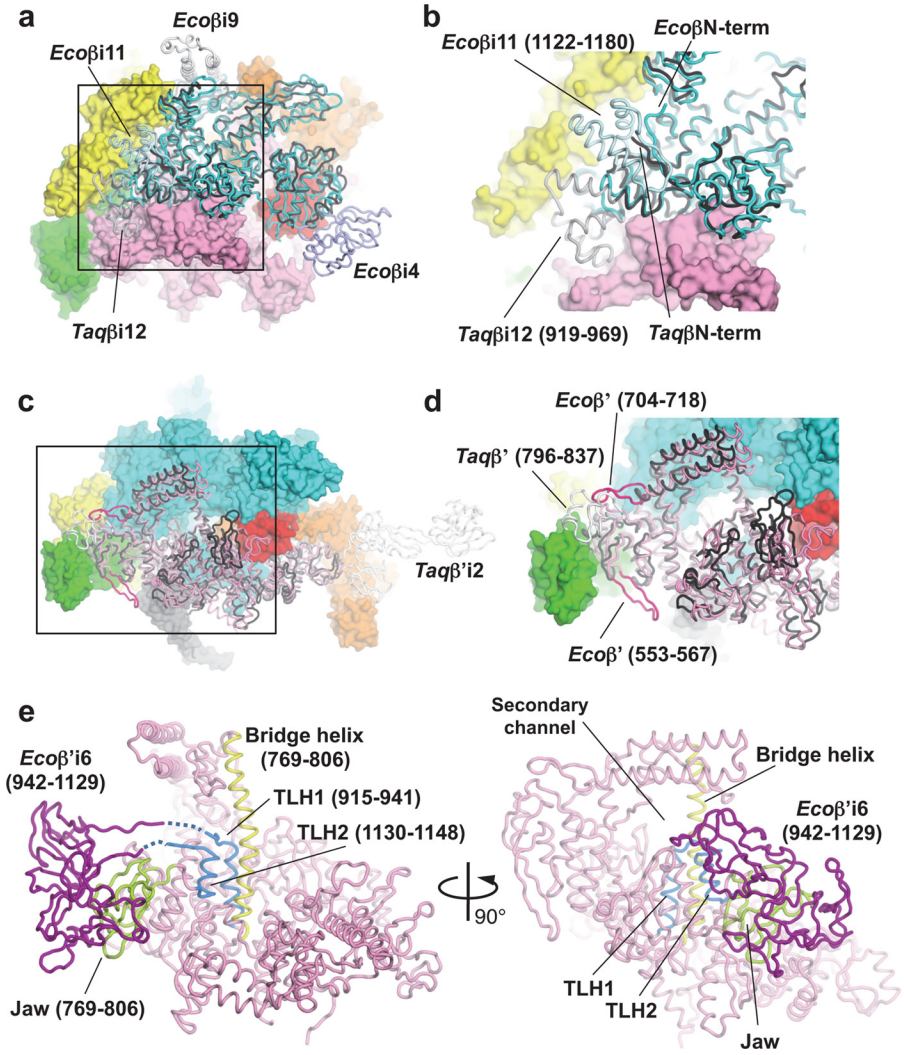

FIGURE 2. **Structure comparisons of the β and β′ subunits of the *E. coli* and *T. aquaticus* RNAPs.** *a*, superposition of *Ecoβ* and *Taqβ* RNAPs. *Ecoβ* (*cyan*) and *Taqβ* (*black*) are shown as α-carbon backbones in addition to the molecular surfaces of other *E. coli* RNAP subunits (α_I, *yellow*; α_{II}, *green*; β′, *pink*; σ^{70}, *orange*; and $\sigma_{1.1}$, *red*). *b*, magnified view of the *boxed* region in *a*. *c*, superposition of *Ecoβ* and *Taqβ* RNAPs. *Ecoβ′* (*pink* and *magenta*), *Taqβ′* (*black* and *white*) are shown as α-carbon backbones in addition to the molecular surfaces of other *E. coli* RNAP subunits (α_I, *yellow*; α_{II}, *green*; β′, *pink*; σ^{70}, *orange*; and $\sigma_{1.1}$, *red*). *d*, magnified view of the *boxed* region in *c*. *e*, the bridge helix (*yellow*), TLH (*light blue*), and jaw (*yellow green*) are highlighted on the α-carbon backbone of the *Ecoβ′* (*pink*) structure. *Ecoβ′*i6 (*purple*) was modeled using the *E. coli* core enzyme model (24).

forms one pincer, called the "clamp," and the β subunit forms the other pincer. The clamp changes its position by swinging between open and closed states (34). Comparison of the *E. coli* RNAP structure with the *Thermus* RNAP structures, including the core enzyme (17, 35), holoenzyme (4, 5), and transcription elongation complex (18), revealed that the *E. coli* RNAP clamp is in a more closed conformation compared with any other RNAP crystal structure solved to date. The gap is narrow (~7.5 Å) between the Cα atoms of σ2 and the tip of the β subunit pincer (residues 371–380) (Fig. 1; see Fig. 4*a* and supplemental Movie S1). The sequences and structures of σ2 and the β subunit pincer are highly conserved in the *E. coli* and *Thermus* RNAPs; therefore, the narrow gap between σ2 and the β subunit pincer observed in the *E. coli* RNAP crystal structure is due to closing of the entire clamp. The $\sigma_{3.2}$ region formed a well ordered loop in the *Thermus thermophilus* holoenzyme (5), but it was disordered in the *Thermus aquaticus* holoenzyme (4). The *E. coli* holoenzyme shows a well ordered $\sigma_{3.2}$ structure (residues 509–519) (see Fig. 4*a* and supplemental Movie S1).

Structural Comparison of E. coli and Thermus RNAPs—The structures of the *E. coli* α subunit dimer and σ^{70} domains σ2 and σ4 have been determined previously (27, 28, 36) and have already been compared with their counterparts in the *Thermus* RNAP (17, 37). However, the *E. coli* RNAP structure from this study enables a direct comparison of the β, β′, and ω subunits between *E. coli* and *Thermus* (Fig. 2).

The entire architecture of the *E. coli* β subunit (*Ecoβ*) can be superimposed on the *T. aquaticus* β subunit (*Taqβ*), with deviations around *Ecoβ*i4 (residues 225–343), *Ecoβ*i9 (residues 938–1042), and *Ecoβ*i11 (residues 1122–1180) (Fig. 2, *a* and *b*). *Ecoβ*i11 comprises three α helices, with a long loop connecting the second and third α helices, and it is located near the β subunit N terminus. The *Ecoβ*i11 structural homolog in the *Thermus* RNAP is βi12 (*Taqβ*i12, residues 919–969), but it is located ~20 Å away from the relative position of *Ecoβ*i11 and

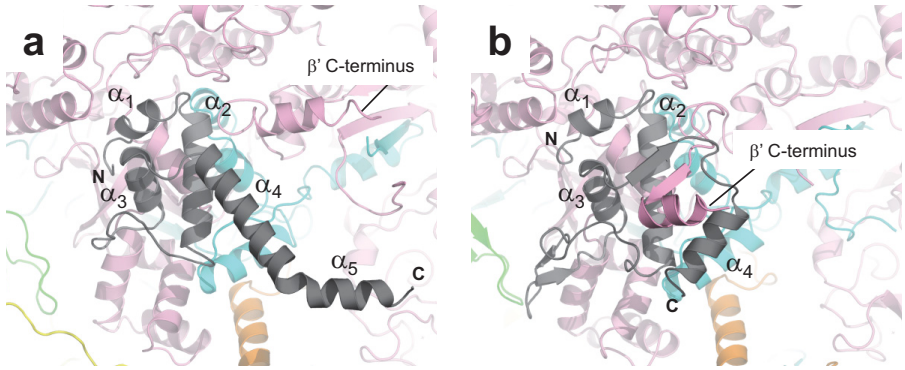

FIGURE 3. **Structure comparisons of the ω subunits of the *E. coli* and *Thermus* RNAPs.** Shown are close-up views of the ω (*gray*) and β′ (*pink*) subunit interactions of the *E. coli* (*a*) and *Thermus* (*b*) RNAPs. The positions of α helices (*E. coli*, α_1–α_5; and *Thermus*, α_1–α_4) of the ω subunits are indicated, and the C termini of the β′ subunits are also indicated.

does not associate with the N-terminal tail of *Taq*β (Fig. 2*b*). The structures of *Eco*β i4 and *Eco*β i9 have been determined and described previously (24).

In the case of the β′ subunits of *E. coli* and *Thermus*, there is structural conservation distributed throughout the entire subunit (Fig. 2*c*). However, *Eco*β′ has several insertions that are not present in *Taq*β′ and vice versa. These insertions include a 13-amino acid insertion between Thr-553 and Thr-567 of *Eco*β′, a 13-amino acid insertion between Glu-704 and Ser-718 of *Eco*β′, and a domain inserted between Arg-796 and Gly-837 of *Taq*β′ (Fig. 2*d*). *Taq*β′ also has a large insert (*Taq*β′i2) between conserved regions A and B (Figs. 1*c* and 2*c*) (24, 38).

The β′ subunit trigger loop/helix (TLH) plays a critical role in the nucleotide addition cycle (39, 40). The front edge of the TLH (residues 930–941 and 1130–1137) is highly flexible, but it becomes a rigid "trigger helix" structure when an incoming nucleotide is present at the active site. The middle of the *E. coli* TLH has a large insert (β′i6, residues 942–1129) that separates the TLH into two regions (TLH1, residues 915–941; and TLH2, residues 1130–1148) (Fig. 2*e*). The edges of TLH1 and TLH2 of *E. coli* RNAP are in loop conformations (residues 930–933 in TLH1 and residues 1133–1138 in TLH2; residues 934–941 and 1130–1132 are disordered). *Eco*β′i6 plays an important role in all stages of transcription, including open complex formation, transcription pausing, and termination, and its location was proposed to be near the β′ subunit jaw (41). However, β′i6 in the *E. coli* holoenzyme structure is completely disordered, without any trace of electron density map, indicating that β′i6 is highly mobile in this crystal structure and possibly in an apo-form holoenzyme.

The β′ subunit bridge helix separates the deep groove of RNAP into a DNA-binding main channel and an NTP entry secondary channel (Figs. 1 and 2*e*) (17). The eukaryotic RNA polymerase II structure shows a straight-form bridge helix (39, 42), whereas the *Thermus* RNAP structures show a bent-form bridge helix (5, 17). Further crystallographic studies of the *Thermus* RNAP complex with the antibiotic streptolydigin (11), as well as a transcription elongation complex (18), have shown that an alternative straight-form bridge helix can exist in the *Thermus* RNAP. Based on these structures, it was proposed that alternate straight-form and bent-form bridge helix conformations are important for the nucleotide additional cycle, including NTP binding and DNA/RNA hybrid translocation (43, 44). The *E. coli* RNAP holoenzyme structure presented here possesses a straight bridge helix (Fig. 2*e*).

Structure and Function of the ω Subunit of E. coli RNAP—The ω subunit of *E. coli* RNAP is composed of five α helices (α_1–α_5) (Fig. 3*a*), and the first three α helices (α_1–α_3) can be overlaid with the first three α helices of the *Thermus* ω subunit (Fig. 3*b*). The folding of the ω subunit N-terminal tail in the *E. coli* RNAP is different in the *Thermus* RNAP structure. The *E. coli* ω subunit C-terminal tail, including α_4 and α_5, is fully extended; the *E. coli* ω subunit makes no interaction with the C-terminal tail of the β′ subunit, in contrast to the *Thermus* RNAP, which has an extensive interaction between the ω subunit and the C-terminal tail of the β′ subunit (supplemental Movie S4).

Functionally, the ω subunit is the least understood subunit, but there is a clear link between the ω subunit and ppGpp-dependent transcription (45, 46). The finding that the ω subunit structure is so different in the *E. coli* and *Thermus* RNAPs may be related to the observation that *E. coli* RNAP can respond to ppGpp only in the presence of the ω subunit (46, 47). Thus, the *E. coli* holoenzyme structure can be used as an ideal system for understanding the relationship between the ω subunit and ppGpp-dependent transcription regulation and may finally reconcile 4 decades of experimental data, especially in understanding the cause of the stringent response and growth control by ppGpp in *E. coli* cells (48–51).

Structure and Function of $\sigma_{1.1}$—Strong and traceable electron density maps of σ^{70} were attainable from $\sigma_{1.2}$ to the C terminus. In one of two *E. coli* RNAP molecules ($RNAP_A$) in the asymmetric unit, the $F_o - F_c$ electron density map calculated using CNS version 1.3 (30) showed rod-like densities for $\sigma_{1.1}$, which is adjacent to $\sigma_{1.2}$. A homology model of *E. coli* $\sigma_{1.1}$, which was constructed by SWISS-MODEL (52) based on the *T. maritima* $\sigma_{1.1}$ NMR structure (6), was placed on the $\sigma_{1.1}$ electron density map, and the positions of three α helices were manually adjusted. An additional α helix (H4) was then built based on a rod-like density next to the third α helix (H3). The $\sigma_{1.1}$ structure was refined in the holoenzyme. The final $\sigma_{1.1}$ structure contains four α helices (residues 6–64), and the electron density of residues from position 65 to $\sigma_{1.2}$ (residue 95) is completely disordered. The higher *B*-factor and weak electron

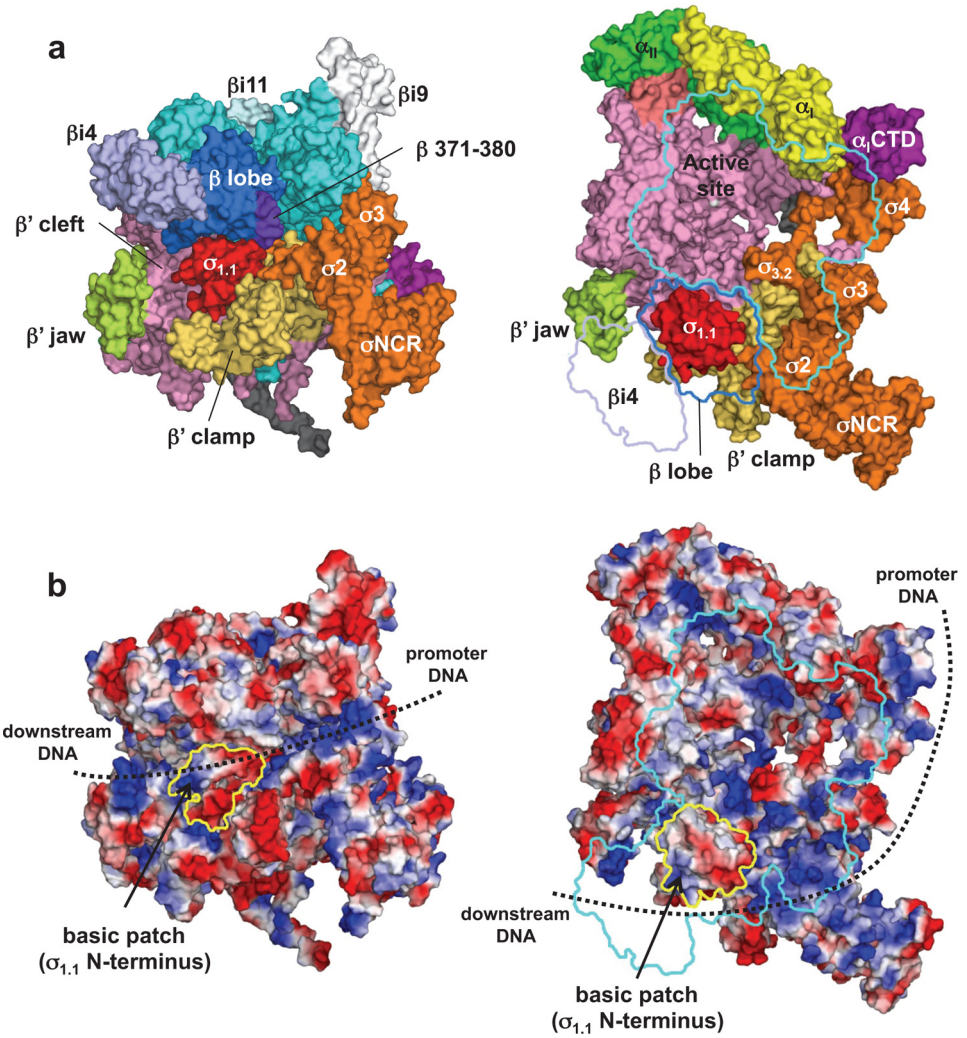

FIGURE 4. **Structure and function of $\sigma_{1.1}$.** *a*, molecular surface of the holoenzyme with $\sigma_{1.1}$. *Left*, front view; *right*, side view. In the right panel, β subunit has been removed and outlined for clarity. *b*, electrostatic distribution of the holoenzyme. *Left*, front view; *right*, side view (orientations are the same as in *a*). Positive electrostatic potential is *blue*, and negative potential is *red*. The positions of $\sigma_{1.1}$ in these views are indicated by *yellow outlines*. A basic patch found at the $\sigma_{1.1}$ N terminus is shown. The potential DNA pathway during open complex formation is shown by *dotted lines*.

density map of $\sigma_{1.1}$ in the *E. coli* holoenzyme structure indicate that $\sigma_{1.1}$ is highly mobile in the holoenzyme.

The structure shows that $\sigma_{1.1}$ is surrounded by σ2, the β lobe, the β' clamp, and the β' cleft (Fig. 4*a* and supplemental Movie S3). The $\sigma_{1.1}$ location in the *E. coli* RNAP crystal structure is consistent with the *E. coli* RNAP model derived from systematic FRET and distance-constrained docking (8).

The acidic residues of $\sigma_{1.1}$ mask the basic residues of the β lobe and β' clamp, and $\sigma_{1.1}$ fits snugly in the DNA-binding main channel of RNAP, thereby preventing access of either double- or single-stranded DNA to the RNAP active site. Therefore, $\sigma_{1.1}$ must disengage from this binding site, or the RNAP clamp must open further (34) to form an open complex.

The structure shows that the three basic residues (Lys-10, Arg-15, and Lys-17) found at the $\sigma_{1.1}$ N terminus are surface-exposed and face the outside of the RNAP main channel (Fig. 4*b* and supplemental Movie S3). These $\sigma_{1.1}$ basic residues, together with other positively charged regions, including σ2, the β lobe, the β' clamp, and the β' jaw, form a continuous path of negative electrostatic potential for promoter DNA and downstream DNA binding. This region may also serve an important role in bending DNA to form the early stage intermediates between the closed and open promoter complexes (7). Although the presence of basic residues at the $\sigma_{1.1}$ N terminus is common in the group 1 σ family, the function of this basic region for transcription has not been tested. This basic region in $\sigma_{1.1}$ could make a contribution to open complex formation.

The α Subunit C-terminal Domain within E. coli RNAP—The C-terminal domain of the α subunit (α-CTD, residues 250–329) is a DNA-binding element and a major target of transcription factors for regulation (22, 53). The two α-CTDs of the RNAP holoenzyme, connected to their N-terminal domains (α-NTD) by linkers (54), can interact independently with transcription factors that bind to DNA 40–100 bp upstream from the transcription start site (55, 56). The structure of an α-CTD in the context of an intact RNAP has not been solved because it is dynamic. In the *E. coli* RNAP structure presented in this study, electron density was visible for only one of the four α-CTDs in the asymmetric unit ($RNAP_A$ α_I). The map enabled

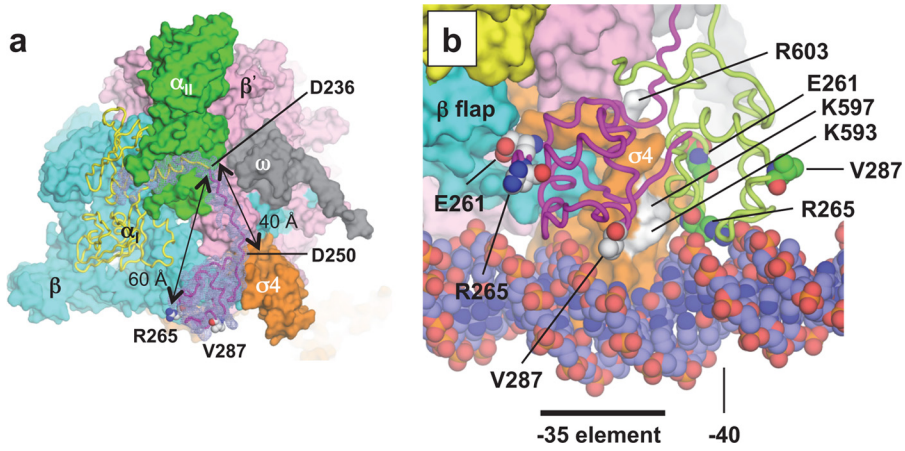

FIGURE 5. **α-CTD structure.** *a*, difference ($F_o - F_c$) electron density map between the holoenzyme and holoenzyme without α_I residues 215–315 (including the last α-NTD α helix, linker, and α-CTD), shown in *mesh*, overlaid on the final holoenzyme structure (α_{II}, β, β′, ω, and σ, surface representation; and α_I, ribbon model). The positions of two key residues for DNA (Arg-265) and transcription factor (Val-287) interaction on the α-CTD are indicated. Glu-261, another key residue of the α-CTD for the σ4 interaction, positioned between the α-CTD and β subunit, cannot be seen from this view. The distance between the α-NTD and Arg-295 and the length of the linker are shown. *b*, superimposed holoenzyme crystal structure (this study) and holoenzyme·catabolite activator protein·DNA model (Protein Data Bank code 3IYD). The two structures were aligned by superimposing their σ4 domains. The α-CTD from the crystal structure (*magenta*) and the model (*light green*) are shown as *ribbons*, and key α-CTD residues are shown as *spheres*. Surface-exposed $\sigma_{4.2}$ residues that are predicted for direct interaction with the α-CTD are colored *white* and are indicated.

a model of the α-CTD to be fitted (57). Furthermore, the map included density of the linker region that allows modeling of the linker (Fig. 5*a*). Arg-265 in the α-CTD is ~60 Å away from the α-NTD (residue 233), with the linker fully stretched and without any secondary structure (Fig. 5*a* and supplemental Movie S2), indicating that the near-maximum length that Arg-265 in the α-CTD can reach DNA from its N-terminal domain is ~60 Å.

Previous biochemical studies suggested that surface-exposed residues in $\sigma_{4.2}$ interact directly with the α-CTD (Fig. 5*b*) (58, 59). Although these residues are partially involved in making the α-CTD·σ4 complex in the *E. coli* RNAP crystal structure, the orientation of the α-CTD relative to σ4 is different compared with the cryo-EM model of the RNAP·catabolite activator protein·DNA complex (60) and the predicted models of the α-CTD·σ4·DNA complex based on biochemical studies (Fig. 5*b* and supplemental Movie S2) (58, 59). The structure of the α-CTD in this holoenzyme structure may be one of several possible conformations of free holoenzyme and would have to rearrange itself for promoter DNA binding.

Concluding Remarks—The crystal structure of the *E. coli* RNAP holoenzyme presented here provides an ideal model for analyzing the functional data that have been generated for over 50 years and for designing future experiments that will uncover the transcription mechanisms. My *E. coli* RNAP structure reveals the molecular features of the α-CTD and $\sigma_{1.1}$ for the first time in the context of an intact bacterial RNAP. Furthermore, I have shown that the *E. coli* RNAP prepared from a co-overexpression vector can generate sufficient quantities of active RNAP for crystallization and high-quality diffraction. This methodology will facilitate the structure determination of the large collection of mutant RNAPs that have been generated for *E. coli* transcription and regulation studies. Finally, because the sequence and antibiotic sensitivity of *E. coli* RNAP are similar to those of pathogen-related RNAPs, including *Mycobacterium tuberculosis* and *Staphylococcus aureus*, *E. coli* RNAP can now be used to readily study RNAP-antibiotic interactions by x-ray crystallography.

Acknowledgments—I thank Scott Pandya, Tomoko Bowser, and Taiki for preparing crystals and Ritwika Basu for synchrotron data collection. I thank the staff at MacCHESS and BCSB for support crystallographic data collection. I thank Seth Darst for providing pACYCDuet-1_Ec_rpoZ. I am grateful to Paul Babitzke, Lucia Rothman-Denes, and Rieko Yajima for critical reading of the manuscript, and Thomas Record, Jr., Amanda Drennan, Pieter de Haseth, Deborah Hinton, and Akira Ishihama for helpful discussions. I thank Nobuyuki Fujita and Shoko Murakami for technical advice for constructing an RNAP-overexpressing vector (pGEMABC). PyMOL was used for preparing figures and calculating the electrostatic potential. This work was initiated in the laboratories of Yasuo Shirakihara at the National Institute of Genetics in Japan and Seth Darst at The Rockefeller University.

REFERENCES

1. Werner, F., and Grohmann, D. (2011) Evolution of multisubunit RNA polymerases in the three domains of life. *Nat. Rev. Microbiol.* **9,** 85–98
2. Gruber, T. M., and Gross, C. A. (2003) Multiple σ subunits and the partitioning of bacterial transcription space. *Annu. Rev. Microbiol.* **57,** 441–466
3. Murakami, K. S., and Darst, S. A. (2003) Bacterial RNA polymerases: the wholo story. *Curr. Opin. Struct. Biol.* **13,** 31–39
4. Murakami, K. S., Masuda, S., and Darst, S. A. (2002) Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. *Science* **296,** 1280–1284
5. Vassylyev, D. G., Sekine, S., Laptenko, O., Lee, J., Vassylyeva, M. N., Borukhov, S., and Yokoyama, S. (2002) Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. *Nature* **417,** 712–719
6. Schwartz, E. C., Shekhtman, A., Dutta, K., Pratt, M. R., Cowburn, D., Darst, S., and Muir, T. W. (2008) A full-length group 1 bacterial σ factor adopts a compact structure incompatible with DNA binding. *Chem. Biol.* **15,** 1091–1103
7. Saecker, R. M., Record, M. T., Jr., and Dehaseth, P. L. (2011) Mechanism of bacterial transcription initiation: RNA polymerase-promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. *J. Mol. Biol.* **412,** 754–771

8. Mekler, V., Kortkhonjia, E., Mukhopadhyay, J., Knight, J., Revyakin, A., Kapanidis, A. N., Niu, W., Ebright, Y. W., Levy, R., and Ebright, R. H. (2002) Structural organization of bacterial RNA polymerase holoenzyme and the RNA polymerase-promoter open complex. *Cell* **108,** 599–614
9. Murakami, K. S., Masuda, S., Campbell, E. A., Muzzin, O., and Darst, S. A. (2002) Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science* **296,** 1285–1290
10. Mukhopadhyay, J., Das, K., Ismail, S., Koppstein, D., Jang, M., Hudson, B., Sarafianos, S., Tuske, S., Patel, J., Jansen, R., Irschik, H., Arnold, E., and Ebright, R. H. (2008) The RNA polymerase "switch region" is a target for inhibitors. *Cell* **135,** 295–307
11. Tuske, S., Sarafianos, S. G., Wang, X., Hudson, B., Sineva, E., Mukhopadhyay, J., Birktoft, J. J., Leroy, O., Ismail, S., Clark, A. D., Jr., Dharia, C., Napoli, A., Laptenko, O., Lee, J., Borukhov, S., Ebright, R. H., and Arnold, E. (2005) Inhibition of bacterial RNA polymerase by streptolydigin: stabilization of a straight-bridge-helix active-center conformation. *Cell* **122,** 541–552
12. Belogurov, G. A., Vassylyeva, M. N., Sevostyanova, A., Appleman, J. R., Xiang, A. X., Lira, R., Webber, S. E., Klyuyev, S., Nudler, E., Artsimovitch, I., and Vassylyev, D. G. (2009) Transcription inactivation through local refolding of the RNA polymerase structure. *Nature* **457,** 332–335
13. Hurwitz, J. (2005) The discovery of RNA polymerase. *J. Biol. Chem.* **280,** 42477–42485
14. Fujita, N., and Ishihama, A. (1996) Reconstitution of RNA polymerase. *Methods Enzymol.* **273,** 121–130
15. Tang, H., Kim, Y., Severinov, K., Goldfarb, A., and Ebright, R. H. (1996) *Escherichia coli* RNA polymerase holoenzyme: rapid reconstitution from recombinant α, β, β', and σ subunits. *Methods Enzymol.* **273,** 130–134
16. Larson, M. H., Landick, R., and Block, S. M. (2011) Single-molecule studies of RNA polymerase: one singular sensation, every little step it takes. *Mol. Cell* **41,** 249–262
17. Zhang, G., Campbell, E. A., Minakhin, L., Richter, C., Severinov, K., and Darst, S. A. (1999) Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell* **98,** 811–824
18. Vassylyev, D. G., Vassylyeva, M. N., Perederina, A., Tahirov, T. H., and Artsimovitch, I. (2007) Structural basis for transcription elongation by bacterial RNA polymerase. *Nature* **448,** 157–162
19. Ho, M. X., Hudson, B. P., Das, K., Arnold, E., and Ebright, R. H. (2009) Structures of RNA polymerase-antibiotic complexes. *Curr. Opin. Struct. Biol.* **19,** 715–723
20. Tupin, A., Gualtieri, M., Leonetti, J. P., and Brodolin, K. (2010) The transcription inhibitor lipiarmycin blocks DNA fitting into the RNA polymerase catalytic site. *EMBO J.* **29,** 2527–2537
21. Vrentas, C. E., Gaal, T., Berkmen, M. B., Rutherford, S. T., Haugen, S. P., Vassylyev, D. G., Ross, W., and Gourse, R. L. (2008) Still looking for the magic spot: the crystallographically defined binding site for ppGpp on RNA polymerase is unlikely to be responsible for rRNA transcription regulation. *J. Mol. Biol.* **377,** 551–564
22. Igarashi, K., and Ishihama, A. (1991) Bipartite functional map of the *E. coli* RNA polymerase α subunit: involvement of the C-terminal region in transcription activation by cAMP-CRP. *Cell* **65,** 1015–1022
23. Doublié, S. (1997) Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol.* **276,** 523–530
24. Opalka, N., Brown, J., Lane, W. J., Twist, K. A., Landick, R., Asturias, F. J., and Darst, S. A. (2010) Complete structural model of *Escherichia coli* RNA polymerase from a hybrid approach. *PLoS Biol.* 10.1371/journal.pbio.1000483
25. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674
26. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326
27. Malhotra, A., Severinova, E., and Darst, S. A. (1996) Crystal structure of a σ^{70} subunit fragment from *E. coli* RNA polymerase. *Cell* **87,** 127–136
28. Blanco, A. G., Canals, A., Bernués, J., Solà, M., and Coll, M. (2011) The structure of a transcription activation subcomplex reveals how σ^{70} is recruited to PhoB promoters. *EMBO J.* **30,** 3776–3785
29. Afonine, P. V., Mustyakimov, M., Grosse-Kunstleve, R. W., Moriarty, N. W., Langan, P., and Adams, P. D. (2010) Joint X-ray and neutron refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 1153–1163
30. Schröder, G. F., Levitt, M., and Brunger, A. T. (2010) Super-resolution biomolecular crystallography with low-resolution data. *Nature* **464,** 1218–1222
31. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60,** 2126–2132
32. Hager, D. A., Jin, D. J., and Burgess, R. R. (1990) Use of Mono Q high-resolution ion-exchange chromatography to obtain highly pure and active *Escherichia coli* RNA polymerase. *Biochemistry* **29,** 7890–7894
33. Mukherjee, K., Nagai, H., Shimamoto, N., and Chatterji, D. (1999) GroEL is involved in activation of *Escherichia coli* RNA polymerase devoid of the ω subunit *in vivo*. *Eur. J. Biochem.* **266,** 228–235
34. Chakraborty, A., Wang, D., Ebright, Y. W., Korlann, Y., Kortkhonjia, E., Kim, T., Chowdhury, S., Wigneshweraraj, S., Irschik, H., Jansen, R., Nixon, B. T., Knight, J., Weiss, S., and Ebright, R. H. (2012) Opening and closing of the bacterial RNA polymerase clamp. *Science* **337,** 591–595
35. Campbell, E. A., Pavlova, O., Zenkin, N., Leon, F., Irschik, H., Jansen, R., Severinov, K., and Darst, S. A. (2005) Structural, functional, and genetic analysis of sorangicin inhibition of bacterial RNA polymerase. *EMBO J.* **24,** 674–682
36. Zhang, G., and Darst, S. A. (1998) Structure of the *Escherichia coli* RNA polymerase α subunit amino-terminal domain. *Science* **281,** 262–266
37. Campbell, E. A., Muzzin, O., Chlenov, M., Sun, J. L., Olson, C. A., Weinman, O., Trester-Zedlitz, M. L., and Darst, S. A. (2002) Structure of the bacterial RNA polymerase promoter specificity σ subunit. *Mol. Cell* **9,** 527–539
38. Chlenov, M., Masuda, S., Murakami, K. S., Nikiforov, V., Darst, S. A., and Mustaev, A. (2005) Structure and function of lineage-specific sequence insertions in the bacterial RNA polymerase β' subunit. *J. Mol. Biol.* **353,** 138–154
39. Wang, D., Bushnell, D. A., Westover, K. D., Kaplan, C. D., and Kornberg, R. D. (2006) Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. *Cell* **127,** 941–954
40. Vassylyev, D. G., Vassylyeva, M. N., Zhang, J., Palangat, M., Artsimovitch, I., and Landick, R. (2007) Structural basis for substrate loading in bacterial RNA polymerase. *Nature* **448,** 163–168
41. Artsimovitch, I., Svetlov, V., Murakami, K. S., and Landick, R. (2003) Co-overexpression of *Escherichia coli* RNA polymerase subunits allows isolation and analysis of mutant enzymes lacking lineage-specific sequence insertions. *J. Biol. Chem.* **278,** 12344–12355
42. Cramer, P., Bushnell, D. A., and Kornberg, R. D. (2001) Structural basis of transcription: RNA polymerase II at 2.8 Å resolution. *Science* **292,** 1863–1876
43. Svetlov, V., and Nudler, E. (2009) Macromolecular micromovements: how RNA polymerase translocates. *Curr. Opin. Struct. Biol.* **19,** 701–707
44. Brueckner, F., Ortiz, J., and Cramer, P. (2009) A movie of the RNA polymerase nucleotide addition cycle. *Curr. Opin. Struct. Biol.* **19,** 294–299
45. Mathew, R., and Chatterji, D. (2006) The evolving story of the ω subunit of bacterial RNA polymerase. *Trends Microbiol.* **14,** 450–455
46. Igarashi, K., Fujita, N., and Ishihama, A. (1989) Promoter selectivity of *Escherichia coli* RNA polymerase: ω factor is responsible for the ppGpp sensitivity. *Nucleic Acids Res.* **17,** 8755–8765
47. Vrentas, C. E., Gaal, T., Ross, W., Ebright, R. H., and Gourse, R. L. (2005) Response of RNA polymerase to ppGpp: requirement for the ω subunit and relief of this requirement by DksA. *Genes Dev.* **19,** 2378–2387
48. Terui, Y., Akiyama, M., Sakamoto, A., Tomitori, H., Yamamoto, K., Ishihama, A., Igarashi, K., and Kashiwagi, K. (2012) Increase in cell viability by polyamines through stimulation of the synthesis of ppGpp regulatory protein and ω protein of RNA polymerase in *Escherichia coli*. *Int. J. Biochem. Cell Biol.* **44,** 412–422
49. Chatterji, D., Ogawa, Y., Shimada, T., and Ishihama, A. (2007) The role of the ω subunit of RNA polymerase in expression of the *relA* gene in *Escherichia coli*. *FEMS Microbiol. Lett.* **267,** 51–55
50. Cashel, M. (1969) The control of ribonucleic acid synthesis in *Escherichia coli.* IV. Relevance of unusual phosphorylated compounds from amino acid-starved stringent strains. *J. Biol. Chem.* **244,** 3133–3141

51. Potrykus, K., and Cashel, M. (2008) (p)ppGpp: still magical? *Annu. Rev. Microbiol.* **62,** 35–51
52. Kiefer, F., Arnold, K., Künzli, M., Bordoli, L., and Schwede, T. (2009) The SWISS-MODEL Repository and associated resources. *Nucleic Acids Res.* **37,** D387–D392
53. Ross, W., Gosink, K. K., Salomon, J., Igarashi, K., Zou, C., Ishihama, A., Severinov, K., and Gourse, R. L. (1993) A third recognition element in bacterial promoters: DNA binding by the α subunit of RNA polymerase. *Science* **262,** 1407–1413
54. Blatter, E. E., Ross, W., Tang, H., Gourse, R. L., and Ebright, R. H. (1994) Domain organization of RNA polymerase α subunit: C-terminal 85 amino acids constitute a domain capable of dimerization and DNA binding. *Cell* **78,** 889–896
55. Murakami, K., Owens, J. T., Belyaeva, T. A., Meares, C. F., Busby, S. J., and Ishihama, A. (1997) Positioning of two α subunit carboxy-terminal domains of RNA polymerase at promoters by two transcription factors. *Proc. Natl. Acad. Sci. U.S.A.* **94,** 11274–11278
56. Tebbutt, J., Rhodius, V. A., Webster, C. L., and Busby, S. J. (2002) Architectural requirements for optimal activation by tandem CRP molecules at a class I CRP-dependent promoter. *FEMS Microbiol. Lett.* **210,** 55–60
57. Jeon, Y. H., Negishi, T., Shirakawa, M., Yamazaki, T., Fujita, N., Ishihama, A., and Kyogoku, Y. (1995) Solution structure of the activator contact domain of the RNA polymerase α subunit. *Science* **270,** 1495–1497
58. Ross, W., Schneider, D. A., Paul, B. J., Mertens, A., and Gourse, R. L. (2003) An intersubunit contact stimulating transcription initiation by *E coli* RNA polymerase: interaction of the α C-terminal domain and σ region 4. *Genes Dev.* **17,** 1293–1307
59. Chen, H., Tang, H., and Ebright, R. H. (2003) Functional interaction between RNA polymerase α subunit C-terminal domain and σ^{70} in UP-element- and activator-dependent transcription. *Mol. Cell* **11,** 1621–1633
60. Hudson, B. P., Quispe, J., Lara-González, S., Kim, Y., Berman, H. M., Arnold, E., Ebright, R. H., and Lawson, C. L. (2009) Three-dimensional EM structure of an intact activator-dependent transcription initiation complex. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 19830–19835