Complete Structural Model of *Escherichia coli* RNA Polymerase from a Hybrid Approach

Natacha Opalka¹, Jesse Brown², William J. Lane³, Kelly-Anne F. Twist¹, Robert Landick⁴, Francisco J. Asturias²*, Seth A. Darst¹*

1 The Rockefeller University, New York, New York, United States of America, 2 Department of Cell Biology, The Scripps Research Institute, La Jolla, California, United States of America, 3 Department of Pathology, Brigham & Women’s Hospital, Boston, Massachusetts, United States of America, 4 Departments of Biochemistry and Bacteriology, University of Wisconsin, Madison, Wisconsin, United States of America

**Abstract**

The *Escherichia coli* transcription system is the best characterized from a biochemical and genetic point of view and has served as a model system. Nevertheless, a molecular understanding of the details of *E. coli* transcription and its regulation, and therefore its full exploitation as a model system, has been hampered by the absence of high-resolution structural information on *E. coli* RNA polymerase (RNAP). We use a combination of approaches, including high-resolution X-ray crystallography, ab initio structural prediction, homology modeling, and single-particle cryo-electron microscopy, to generate complete atomic models of *E. coli* core RNAP and an *E. coli* RNAP ternary elongation complex. The detailed and comprehensive structural descriptions can be used to help interpret previous biochemical and genetic data in a new light and provide a structural framework for designing experiments to understand the function of the *E. coli* lineage-specific insertions and their role in the *E. coli* transcription program.

**Introduction**

RNA in all cellular organisms is synthesized by a complex molecular machine, the DNA-dependent RNA polymerase (RNAP). In bacteria, the catalytically competent core RNAP (subunit composition α₂ββ′σ₀) has a molecular mass of ~400 kDa. Evolutionary relationships for each of the bacterial core subunits have been identified between all organisms from bacteria to man [1–3]. These relationships are particularly strong between the two largest subunits, β' and β, which contain colinearly arranged segments of conserved sequence (Figure 1) [3]. These conserved segments are separated by relatively nonconserved spacer regions in which large, lineage-specific sequence insertions and their role in the *E. coli* transcription program.**
contacts in the helical crystals. Strong electron density for Ec β9 was present in the cryo-EM reconstruction, but weak density for Ec β4 and Ec β6 indicated these domains were flexible in the context of the helical crystals [13]. Most previous EM reconstructions of various forms of Eco RNAP have not revealed information concerning the lineage-specific inserts (for instance, see [14]). A recent 20 Å-resolution, negative-stain EM reconstruction of an activator-dependent transcription initiation complex containing Eco RNAP [15] allowed the positioning of the Ec β6 crystal structure [10], but the lack of structural information on the other Eco lineage-specific inserts prevented the detailed interpretation of additional densities present in the reconstruction [15].

In this study, we used a combination of structural approaches to generate a complete molecular model of Eco core RNAP. We determined two new high-resolution X-ray crystal structures of Eco RNAP β subunit fragments that include Eco β4 and β9 and used an ab initio method to predict the structure of the small Eco β11 [16]. The three available X-ray crystal structures of Eco RNAP fragments (the two structures determined herein and the structure of Eco β6 [10]) and the predicted structure of Eco β11 were incorporated into a homology model of Eco core RNAP. Finally, we used cryo-EM imaging combined with single-particle image analysis to obtain a low-resolution structure of the solution conformation of Eco core RNAP in which densities corresponding to lineage-specific insertions could be clearly identified. Flexible-fitting of the Eco RNAP homology model into cryo-EM densities generated a complete molecular model of Eco core RNAP and an Eco RNAP ternary elongation complex (TEC).

Results

Crystal Structure of Eco RNAP βflap-β9

The lineage-specific insert β6 (previously named β dispensable region 2, or BDR2, or SI2 in the literature [13,18,22,23]) is located between bacterial shared regions β8β6 and β8β7 (using the bacterial RNAP common region nomenclature of Lane et al. [3]) in the β2 domain (Figure 1) [5,19], was predicted to comprise from one to six tandem repeats of a structural motif termed the β-β module 2 (BBM2) [4]. The β4 of Acidobacteria, Mollicutes, and Proteobacteria (including E. coli) was predicted to comprise two tandem BBM2 repeats [3]. Eco β4 comprises β residues 225–343 (Figure 2A).

We prepared a construct comprising the Eco β2 domain including β6 inserted within it (Eco β residues 152–443, hereafter called Ec β2-β4). After reductive methylation [20], the protein formed crystals that diffracted X-rays to 1.6 Å-resolution (Table 1). The structure was solved by single-anomalous dispersion using a dataset collected from crystals of selenomethionyl-substituted protein [21] and refined to an R/Refree of 0.209/0.229 at 1.6 Å-resolution (Table 1, Figures 2, S1).

As expected, the Eco β2 (Eco β residues 151–224 and 344–445) and the Thermus β2 (Taq or Tth β residues 138–231) domains have similar overall structures (Figure S2). A superimposition of the two domains over 100 residues (excluding flexible loops connecting secondary structural elements) yields a root-mean-square deviation in x-carbon positions of 1.68 Å. Significant differences in the structures include: (i) the loop connecting the first two β-strands of the β2 domain, where Ecol has a 5-residue insertion (Eco β residues 164–168, disordered in our structure), and (ii) the loop connecting the last two β-helices of the β2 domain, which includes a 7-residue insertion present in Taq β (Taq β residues 293–299; Figures 2A, S2).

The β4 domain is inserted at the surface of the β2 domain distal to the connection with the RNAP (Figure 2B). A 3-residue segment of Taq β (Taq β 212–214) is replaced by the 119-residue Eco β4 (Figure 2A). The Eco β4 folds into a compact, cylinder-shaped domain about 22 Å in diameter and about 50 Å in length (Figures 2B, 2C). The compact domain is connected to the β2 domain by two short connector loops (Eco β 225–226 and 337–345). The β4 domain packs against β2, resulting in the burial of a modest 618 Å² of surface area. As predicted [4], the Eco β4 includes two tandem BBM2 motifs (Figure 2A, 2C).

Crystal Structure of Eco RNAP βflap-β9

The lineage-specific insert β9 (previously named β dispensable region 2, or BDR2, or SI2 in the literature [13,18,22,23]) is located between bacterial shared regions β13 and β14 [3] at the base of the flap domain (Figure 1) [5,19]. The β9 is found in Acidobacteria, Aquificae, Bacteroidetes, Chloramydiae, Chlorobi, Planctomycetes, Proteobacteria (including E. coli), and Nitrospirae [3]. Eco β9 comprises β residues 938–1042 (Figure 3A).

A construct comprising the Eco flap domain (Eco β 831–1057), including β9, was crystallized as a complex with bacteriophage T4 gp33 (K.-A.F.T., P. Deighan, S. Nechaev, A. Hochschild, E.P. Geiduscheck, S.A.D., in preparation). The structure was solved by a combination of molecular replacement (using the Taq flap domain as a search model) and single-anomalous dispersion using data collected from selenomethionyl-substituted protein (Table S1, Figure S3) [21]. The complete structure was refined to an R/Refree of 0.264/0.291 at 3.0 Å-resolution. T4 gp33 interacts primarily with the flap-tip and does not make any interactions with β9. These and further details of the complex with T4 gp33 will be described elsewhere (K.-A.F.T., P. Deighan, S. Nechaev, A. Hochschild, E.P. Geiduscheck, S.A.D., in preparation).

The β9 domain is inserted at the base of the flap domain, near the C-terminal connection of the flap with the rest of the RNAP and distal to the flap-tip (Figure 3B). A 6-residue segment of Taq β (Taq β 809–814) is replaced by the 105-residue Eco β9 (Figure 3A). The Eco β9 comprises two long, parallel β-helices of 38 and 32
residues (Eco β 943–980 and 1006–1037, respectively) with a short, hook-like connecting segment (residues 981–1005) at the end distal to the flap (Figure 3B), forming an apparently rigid structure reminiscent of a hook-and-ladder that extends nearly 65 Å out from the flap domain. The β9 is connected to the flap domain by two connector loops (Eco β 938–942 and 1038–142) but makes minimal interactions with the flap itself. The structure does not appear to conform to the β-β′ module 1 motif (BBM1, similar to the BBM2 motif, Figure 2C) predicted for β9 [4]. The 105-residue Eco β9 is at the lower end of the size range for β9 sequences, which ranges from 105 residues in some Proteobacteria to 143 residues in some Bacteriodetes. An alignment of 307 non-redundant β9 sequences (see Dataset S1) reveals that the two long, ladder α-helices do not harbor insertions; all of the insertions occur in the hook-like connector at the distal end of β9 (Figure 3A). Therefore, we conclude that β9 has a conserved core structure with the two ladder α-helices of conserved length.

Cryo-EM Reconstruction of Eco RNAP

We generated a single-particle cryo-EM (spEM) reconstruction of Eco RNAP by analyzing ~42,000 images of Eco RNAP particles preserved in vitreous ice (Figures 4A, S4–S6). Initial image orientation parameters were determined using a 35 Å-resolution RNAP model based on the Tag core RNAP X-ray structure [5]. Final refinement of image orientation parameters by projection matching yielded a structure of Eco RNAP with a 0.5 Fourier-shell cutoff resolution of ~11.2 Å (Figure S4). Nevertheless, information beyond about 14 Å resolution was very weak, and so the figures and analysis described herein were performed on a low-pass Fourier-filtered map [24,25]. Although the cryo-EM grids were prepared with samples of Eco RNAP holoenzyme (core RNAP plus the promoter-specificity σ70 subunit), the σ70 subunit apparently dissociated during grid preparation as density corresponding to σ70 was completely absent. Dissociation during cryo-EM sample preparation has been noted for other macromolecular complexes [26] and is also consistent with reports of dissociation constants for the σ70/core RNAP complex as high as 200–300 nM (the RNAP concentration used here was about 200 nM). The spEM reconstruction showed Eco core RNAP in a conformation similar to that observed in Thermus X-ray structures but with clear density corresponding to β4, β11, and β96 (Figures 4A, S5, S6).

Molecular Model of the Complete Eco Core RNAP

In order to interpret the spEM map of Eco core RNAP, we generated a homology model of Eco core RNAP using the core component of the T. thermophilus (Tth) RNAP holoenzyme structure.
Figure 2. Sequence and structure of Eco RNAP β2-βi4. (A) Sequence alignment comparing Eco RNAP β2-βi4 with the corresponding region of Taq (which lacks βi4). Shaded residues are identical between the two sequences. The secondary structures are indicated directly above (for Eco) and below (for Taq) the sequences; filled rectangles denote α-helices, open rectangles denote β-strands, the dashed lines denote disordered regions. The number scale above the Eco secondary structure corresponds to the Eco β subunit sequence. Above the number scale, black lines denote the sequence regions common to all bacterial RNAPs [3]. The yellow and orange lines denote the two BBM2 motifs [4]. The extent of the common β2 domain (thick cyan line) and the lineage-specific insert βi4 (thick green line) is indicated at the top. (B) Ribbon diagram of Eco β2-βi4 (β2 domain, cyan; βi4, green). A disordered loop (Eco β 161–169) is denoted by small spheres. The view corresponds to the reference view of Taq core RNAP (lower left, β-side view), shown as a backbone worm and color-coded as follows: αl, αIl, ω, gray; β, light pink; βi, light cyan, except the β2 domain is colored cyan and labeled. (C) Ribbon diagram of Eco βi4 (same view as B). The tandem BBM2 motifs predicted by Iyer et al. [4] are color-coded as in (A) (BBM2a, yellow; BBM2b, orange). doi:10.1371/journal.pbio.1000483.g002
The crystal structures of Eco RNAP (PDB ID 1IW7) [7] as a template. The locations of the Eco lineage-specific insertions βi4, βi9, βi11, and βi16 (absent in Thermus) were left as gaps in the Eco sequences. Thermus-specific inserts βi2 and βi2 (Figure 1) were also removed from the structural template. The crystal structures of Eco β2-βi4 (Figure 2B) and βi11-βi16 (Figure 3B) were spliced into the resulting homology model by superimposition of the overlapping β2 and βi1 domains, respectively. At this stage, the Eco RNAP model was readily fit manually into the spEM map. The spEM map contained clear density corresponding to βi4, but density for βi9 was absent. Density for the Ω subunit as well as the C-terminal helix of β were also absent. In addition, extra density not accounted for by the homology model was present for βi11 and βi16. An ab initio predicted structure of the short βi11 (see below) was placed into the corresponding density to fill in the gap in the Eco β sequence between 1121 and 1181. The crystal structure of Eco βi6 (PDB ID 2AUK) [10] was readily fit manually into excess density in the vicinity of its insertion point in β. Two criteria were used to determine the orientation of βi6 with respect to the rest of the RNAP. First, although βi6 comprises a tandem repeat of two SBHM domains, the C-terminal SBHM domain (SBHMb) [10] harbors larger insertions between the core SBHM β-strands, making βi6 asymmetric in shape. The asymmetry is clearly seen in the spEM density as well (see Figure 4A, top view). Moreover, only one orientation of βi6 allows connection to the gap in the Eco β sequence (between residues 940 and 1132) without severe distortion. The positioned βi6 was readily connected to the open (unfolded) trigger-loop (TL) conformation of the model.

Flexible-fitting of the final Eco RNAP model (excluding Ω, the C-terminal 41 residues of β, and βi9) into the spEM map was performed using YUP.SCX [27], resulting in a superb fit of the conserved RNAP as well as of the lineage-specific inserts (excluding βi9; Figures 4A, S5, S6). In order to position βi9 in the context of the entire RNAP structure, we used our previously determined helical cryo-EM map of Eco core RNAP (hEM) and fit of the Taq core RNAP X-ray crystal structure [13] since the hEM map contains strong density for βi9. The flap portion (excluding the flexible flap-tip) of the Eco βi16-βi19 crystal structure (Figure 3B) was superimposed on the Taq βi16 domain in the context of the Taq RNAP fit into the hEM density. The resulting position of βi9 did not correspond to the hEM density (light orange, βi9 in Figure 4B) but was fit into the density by a rotation of about 35° (orange, βi9 in Figure 4B). This positioning of βi9 is consistent with the location of positive difference density observed in the context of the helical crystals due to a 234-residue insertion between Eco β residues 998 and 999 (red dot, Figure 4B). The Eco core RNAP model was completed by adding back the C-terminal segment of βi6 as well as Ω (in accordance with the Thermus RNAP structures).

The Eco core RNAP model was then used as the basis for generating a homology model of an Eco TEC, using the Tth TEC crystal structure (open TL conformation, PDB ID 2D05) [8]. For both models, the lineage-specific inserts (βi4, βi9, βi11, βi16 for Eco βi2 and βi12 for Tth) were removed. The nucleic acids present in the Tth crystal structure were fixed during the modeling. The Eco lineage-specific inserts were added back to the resulting TEC model (according to their positions in the Eco core RNAP model), and missing portions of the nucleic acids (the upstream double-stranded DNA, and the nontemplate strand of the DNA within the transcription bubble) were modeled according to Korzhova et al. [28].

Discussion

In this work, two new X-ray crystal structures (Eco β2-βi4, Figure 2; Eco βi16-βi9, Figure 3) and an ab initio predicted structure (Eco βi11, see below), combined with a previously determined X-ray crystal structure of Eco βi6 [10], provide high-resolution structural descriptions of each of the lineage-specific sequence insertions found in the highly biochemically and genetically characterized Eco RNAP [3]. In addition, a new 15 Å-resolution cryo-EM single-particle reconstruction of Eco RNAP (Figures 4A, S4–S6) reveals clear electron density for βi4, βi11, and βi16, while a previously determined cryo-EM reconstruction of Eco core RNAP from helical crystals contains strong electron density for βi9 [13,23]. The combination of these structural data provides the basis for a detailed and complete atomic model of Eco RNAP and an Eco core RNAP TEC.

The large β and β subunits comprise regions of sequence shared among all bacterial RNAPs [3]. These shared regions, which make up 63% of the Eco β and 67% of the Eco β sequence, are expected to have nearly identical structure among all bacterial RNAPs. The Ω subunits are also highly homologous [5,29]. Thus, most of the Eco RNAP structure is expected to be highly similar, if not identical, to the Thermus RNAP structures. The unique contribution of this work is the high-resolution structural information on the Eco lineage-specific inserts βi4, βi9, and βi11, as well as the detailed structural model of all of the lineage-specific

| Table 1. Crystallographic statistics for Eco RNAP β2-βi4 crystals. |
|---------------------------------------------------------------|
| **Se1** | **Se2** |
| **Data collection** | | |
| Space group | P2,2,2 | P2,2,2 |
| Cell dimensions | | |
| ɑ,  β,  γ (Å) | 106.28, 51.84, 61.77 | 106.31, 52.04, 61.83 |
| α, β, γ (%) | 90, 90, 90 | 90, 90, 90 |
| Wavelength | 0.9785 | 0.9919 |
| Resolution (Å) | 25.0–1.90 (1.97–1.90) | 25.0–1.60 (1.64–1.60) |
| R<sub>work</sub> | 0.081 (0.596) | 0.0690 (0.416) |
| R<sub>ref</sub> | 11.0 (2.7) | 40 (5.1) |
| Completeness (%) | 94.1 (87.1) | 98.5 (94.0) |
| Redundancy | 2.6 (2.4) | 7.0 (6.5) |
| **Refinement** | | |
| Resolution (Å) | 25.0–1.60 | 25.0–1.60 |
| No. reflections | 42,737 | 42,737 |
| R<sub>work</sub>/R<sub>ref</sub> | 0.209/0.229 | 0.209/0.229 |
| No. atoms | | |
| Protein | 2,345 | 2,345 |
| Water | 386 | 386 |
| β-factors | | |
| Protein | 14.51 | 14.51 |
| Water | 24.58 | 24.58 |
| R.m.s deviations | | |
| Bond lengths (Å) | 0.008 | 0.008 |
| Bond angles (°) | 1.134 | 1.134 |

*Scaling statistics for Se1 dataset calculated without combining anomalous pairs.
 doi:10.1371/journal.pbio.1000483.t001
inserts in the context of the entire RNAP and a TEC. The following discussion therefore focuses on the Eco lineage-specific inserts and insights into their role in RNAP function provided by our new structural information.

**βi4**

RNAPs harboring deletions or insertions within βi4 support cell growth and retain basic in vitro transcription function, leading to its designation as “dispensable region I” of the β subunit [17]. Nevertheless, careful studies of a nearly precise βi4 deletion (deletion of Eco β 226–350) revealed defects [18]. The purified Δβi4-RNAP showed only very mild defects, or no defects at all, in a number of in vitro tests [17,18]. In vivo, however, the Δβi4-RNAP was unable to support cell growth at 42°C and could only support slow growth at 30°C.

In our model of the Eco TEC, βi4 extends out from the β2 domain roughly in the direction of the downstream double-stranded DNA (Figure 5). However, βi4 is unlikely to interact

---

**Figure 3. Sequence and structure of Eco RNAP βflap-βi9.** (A) Sequence alignment comparing the sequence context of Eco RNAP βi9 with the corresponding region of Taq (which lacks βi9). Shaded residues are identical between the two sequences. The secondary structure for Eco is indicated directly above the sequence; filled rectangles denote α-helices, open rectangles denote β-strands. The number scale above the Eco secondary structure corresponds to the Eco β subunit sequence. Above the number scale, black lines denote the sequence regions common to all bacterial RNAPs [3]. Gaps in the βi9 sequence with numbers above denote the location and residue length of insertions in an alignment of 307 non-redundant βi9 sequences (see Supporting Information). The extent of the common βflap domain (thick cyan line) and the lineage-specific insert βi9 (thick orange line) is indicated at the top. (B) Two orthogonal views of Eco βflap-βi9 (βflap, cyan; βi9, orange). The views correspond to the reference views of Taq core RNAP (left, bottom view; right, front view), shown as a backbone worm and color-coded as follows: αI, αII, ε, gray; β’, light pink; β, light cyan, except the βflap domain is colored cyan and labeled. doi:10.1371/journal.pbio.1000483.g003
directly with the downstream DNA to form part of an extended DNA binding channel since βi4 tilts away from the DNA, creating a roughly 15 Å gap between itself and the DNA. Moreover, the solvent-exposed surface of βi4, including the entire surface facing the DNA, is highly acidic (Figure 5, front view), except for a “neutral patch” that arises from three conserved residues, Eco β R368, R372, and R375 (Figure 5, top view). These positions are conserved as basic residues (either R or K) in 90%, 91%, and 91% of the sequences, respectively, in an alignment of 316 non-redundant βi4 sequences (containing only “Eco-like” βi4 sequences comprising two BBM2 domains; see Dataset S2) and may comprise an interaction determinant for an as yet unidentified regulatory factor.

The bacteriophage T4 Aε protein interacts with the host Eco RNAP [30] and causes premature transcription termination on Eco DNA while allowing Eco RNAP-mediated transcription of phage DNA containing 5-hydroxymethylcytosine [31]. Eco paf mutants (prevent Aε function) have been mapped to the i9, a region shared among all bacterial RNAPs (Figure 2A) [3]. In our structural model of the Eco RNAP TEC, βR368 and βP357 lie within a structural feature that sits at the entrance of the main RNAP active site channel, inside the “V” formed by the upstream and downstream DNA of the TEC (Figure 5, channel and front views). These residues are not near any nucleic acids in the TEC (the closest approach is for the backbone carbonyl of βP357, which is 15 Å away from the non-template DNA phosphate backbone at the -10 position) but could comprise part of an Aε binding determinant on the RNAP [17]. The 19 kDa Aε protein bound in this vicinity (Figure 5, channel and front views) would be well positioned to distinguish the presence of cytosine or 5-hydroxymethylcytosine in either the downstream double-stranded DNA (where the 5-hydroxymethyl moiety would be exposed in the major groove) or the single-stranded non-template DNA in the transcription bubble.

βi9 RNAPs harboring deletions or insertions within βi9 support cell growth and retain in vitro transcription function, leading to its designation as “dispensable region II” of the β subunit [17,22,29,33]. Nevertheless, careful studies of a precise βi9 deletion (deletion of Eco β 938–1040) revealed defects [18]. The purified Δβi9-RNAP showed only very mild defects, or no defects at all, in a number of in vitro tests [18]. The βi9 contains the epitope for the PYN-6 monoclonal antibody and, consistent with in vitro tests showing little effect of deleting βi9 on normal RNAP function, RNAP can be immobilized using the PYN-6 antibody but remains active for in vitro transcription [22]. In vivo, however,
the Δβ9-RNAP was unable to support cell growth in minimal media [18].

Our crystal structure of the Eco βflap-β9 suggests that β9 is attached to the flap via flexible linkers and does not make a significant, stable interaction with the flap (Figure 3B), suggesting that β9 is highly flexible in its orientation with respect to the flap. Indeed, the position of β9 in the βflap-β9 crystal structure appears to be determined by packing interactions with neighboring, symmetry-related molecules. In keeping with this, there is no density for β9 in the spEM reconstruction (Figures 4A, S5, S6). However, in our previous hEM reconstruction of Eco RNAP, strong density consistent with β9 was observed, and this density was shown to correspond to β9 through a helical reconstruction of a mutant RNAP harboring a large insertion between positions 998 and 999 [23]. In the helical crystals, the packing of a neighboring, symmetry-related RNAP molecule restricts the range of positions available to β9, allowing its visualization (Figure 4B). Fitting β9 into the corresponding density in the hEM reconstruction required a large change in the position of β9 with respect to the flap, but the final model fits very well into the density and is also consistent with the EM localization results [23], which were not used as a constraint in the fitting (Figure 4B). This model for the position of β9 in the context of the entire RNAP is presented as an example of a particular orientation that is possible for β9 (since it was observed in the helical crystals), but the evidence indicates that β9 does not adopt a particular conformation with respect to the RNAP but can access a wide range of positions (Figure 6).

Figure 6. Orientational flexibility of β9. Bottom view of the Eco RNAP model. The RNAP is shown as a molecular surface (α1, αII, ω, grey; β, light cyan, except β9 is green and β11 is magenta; β′, light pink) except for β9, which is shown as a backbone worm. The modeled position of β9 (see Figure 4B) is colored orange. Selected alternative orientations accessible to β9 are colored light orange. The potential reach of β9 maps out roughly a hemisphere with a radius of 65 Å.
doi:10.1371/journal.pbio.1000483.g006
The modeled position of βi9 is not near any nucleic acids in the TEC or in the open promoter complex [34]. Moreover, the solvent-exposed surface of βi9 is primarily acidic (Figure S7). Interestingly, an alignment of 307 non-redundant βi9 sequences (see Dataset S1) reveals that conserved, solvent-exposed residues are all displayed on the back face of the “ladder,” opposite the “hook” (Figure S7). Conserved features of this face comprise charged residues D959 (conserved as D or E in 97% of the sequences), E962 (D/E, 95%), R974 (K/R, 99%), K1032 (K/R, 95%), and K1035 (K/R, 94%), and one conserved hydrophilic residue, I966. These features suggest that this face of the ladder may serve as an interaction determinant for as yet unidentified regulatory factors. D959 and K1032 participate in a clearly conserved salt bridge.

βi11

The lineage-specific insert βi11 is located between bacterial shared regions βi14 and βi15 (Figures 1, 7A) [3]. The βi11 is found in Acidobacteriaceae, Aquificae, and Proteobacteria (including E. coli) [3]. In each bacterial species where it is found, βi11 has a length ranging from 54–69 residues. Comparing Taq with E. coli, a 5-residue segment of Tag β (Tag β 895–899) is replaced by the 59-residue E. coli βi11, comprising E. coli β residues 1122–1180 (Figure 7A).

Although a construct corresponding to E. coli RNAP βi11 overexpressed and was well behaved, we were unable to obtain crystals suitable for X-ray analysis. The Robetta server (http://roberta.bakerlab.org/) provided an ab initio predicted structure of this short, 59-residue fragment (Figure S8) that is consistent with a number of observations from our structural and sequence analyses:

(i) The overall predicted structure of βi11 fits well into the corresponding spEM density (Figure 7B, right).

(ii) The termini of the predicted βi11 structure could be readily connected to the corresponding gap in the E. coli RNAP β structure with only minor modifications.

(iii) In an alignment of 310 non-redundant βi11 sequences (see Dataset S3), insertions and gaps occur in locations consistent with the predicted structure (i.e. in loops connecting secondary structural elements and away from the RNAP; Figure S8).

(iv) Analysis of the βi11 sequence alignment reveals that most of the conserved residues are hydrophobic in nature and are buried in the hydrophobic core of the βi11 fold (Figure S8C). Two conserved, solvent-accessible polar residues (R1142 and D1166) form an apparently conserved salt-bridge that may stabilize the structure (Figure S8C).

The βi11 was only recently recognized as a distinct, lineage-specific insertion [3,4]. To our knowledge, no information on the effects of deletions or mutations in this region is available. Inspection of the spEM map and the aligned X-ray structure of Taq core RNAP in the region of the β subunit between shared regions βi14 and βi16 revealed a clear discrepancy that corresponds to Tag βi12 (Figure 7B). In our E. coli RNAP model, the Tag βi12 was removed and the resulting gap was connected by the loop corresponding to E. coli β residues 1200–1207. The predicted structure of E. coli βi11 (Figure S8) was then spliced between E. coli β residues 1121 and 1181 and oriented to fit into the EM density, resulting in a good fit. The resulting location of E. coli βi11 clashed with the position of the β-subunit N-terminus, which was redirected to relieve the clash (Figure 7B).

βi6

While the large E. coli lineage-specific insertions βi4 and βi9 appear to play only peripheral roles in RNAP function, and the complete deletion of either one results in relatively minor growth defects [18], βi6 plays a more important role in E. coli RNAP function. Complete deletion, or even partial deletion, of βi6 is not viable [18,35]. Complete deletion causes a severe defect in RNAP assembly, both in vivo and in vitro [18,35], but the in vivo-assembled Δβi6-RNAP can be obtained from cells simultaneously overexpressing the other RNAP subunits [18], and partial deletions of βi6 can be assembled in vitro [35]. Biochemical studies of enzymes with complete or partial βi6 deletions reveal a number of severe defects. The Δβi6-RNAP forms dramatically destabilized open promoter complexes [10]. RNAPs harboring partial deletions in βi6 are defective in transcript cleavage and have a dramatically reduced transcript elongation rate at subsaturating NTP concentrations [35]. Antibody binding to epitopes within βi6 inhibit transcription as well as intrinsic transcript cleavage [35,36].

The βi6 plays a central role in the pausing/termination behavior of elongating E. coli RNAP [18,35]. Full or partial deletions in βi6 result in RNAPs with dramatically altered pausing behavior [18,35]. A genetic screen for termination-altering mutants in E. coli RNAP uncovered 10 positions scattered throughout βi6 [37].

These profound effects of βi6 on E. coli RNAP function are likely due to its insertion in the middle of a critical and highly conserved structural feature of the RNAP, the so-called “trigger-loop” (TL), which connects two highly conserved α-helices (TL-helices 1 and 2, TLH1 and TLH2; Figures 1, 8). The TLHs, in turn, interact with another central structural element, the bridge-helix (BH; Figure 8B). The TL tends to be unstructured (open) in RNAP and in the substrate-free TEC but is found in a structured conformation (closed) where it makes many direct contacts with the incoming NTP substrate in the TEC [38,39]. The TL has been proposed to cycle between open and closed conformations at each nucleotide addition step to promote rNTP substrate recognition, enzyme fidelity, and possibly catalysis [38–42].

Microcin J25 (MccJ25) is a bactericidal 21-residue peptide that inhibits transcription by binding bacterial RNAP within the secondary channel [43–46]. Based on saturation mutagenesis of E. coli RNA polymerase (the gene encoding the RNAP β subunit), MccJ25 does not contact βi6; most amino acid substitutions that yield strong resistance against MccJ25 lie in the BH and the TL [43,44,46]. Nevertheless, a deletion of βi6 perturbs the effects of MccJ25 [46], likely through the effects of the βi6 deletion on the TL conformation.

Our positioning of βi6 in the spEM density (Figures 4, S3, S6) and its connections with the open TL conformation (Figure 8B) are similar to the results of Hudson et al. [15]. The βi6 sits outside the RNAP active site channel and makes extensive interactions with the β-jaw (Figure 8B). The N-terminal SBHM domain of βi6 (SBHMa) faces the secondary channel, consistent with the results of crosslinks mapped from backtracked TECs (in which the 5′-end of the RNA transcript is extruded out the secondary channel) between analogs incorporated into the RNA 3′-end and the N-terminal region of βi6 [28]. SBHMb faces the downstream double-stranded DNA-binding channel (Figures 5, 8) but does not contact the DNA; the closest approach between the DNA and βi6 is 16 Å (between βD1073 and the non-template strand backbone phosphate at +14). Moreover, βi6 is highly

Structural Model of E. coli RNA Polymerase
acidic over its entire solvent-exposed surface, including the region facing the downstream double-stranded DNA (Figure 5, front view).

Although βi6 connects readily to the open conformation of the TL via extended linkers (Figure 8B), modeling suggests it would not be able to connect with the closed TL conformation in the modeled position, a conclusion also reached by Hudson et al. [15]. Since the folding of the TL is required for interactions between highly conserved TL-residues and the incoming nucleotide substrate [19,38,39], it is likely that the position of βi6 must change to accommodate the folded TL conformation at each nucleotide addition step of the transcription cycle.

During bacteriophage T7 infection, the Eco RNAP β' subunit is phosphorylated by the phage-encoded kinase Gp0.7 [47], and the site of phosphorylation has been identified as a single amino acid in βi6, T1068 (Figures 5, 8) [48]. Phosphorylation at this site appears to affect pausing, as well as ρ-dependent termination behavior, of Eco RNAP [48]. This site is in the βi6 loop that makes the closest approach to the downstream DNA, but as discussed above, this region is nevertheless not in close contact

---

**Figure 7. Sequence and structural context of Eco RNAP βi11 and Taq βi12.** (A) Sequence alignment comparing the sequence context of Eco RNAP βi11 with the corresponding region of Taq (which lacks βi11 but harbors βi12) [3]. Shaded residues are identical between the two sequences. The experimentally determined secondary structure for Taq is indicated directly below the sequence; filled rectangles denote α-helices, open rectangles denote β-strands. The number scale above the Eco secondary structure corresponds to the Eco β subunit sequence. Above the number scale, black lines denote the sequence regions common to all bacterial RNAPs [3]. The extent of Eco βi11 and Taq βi12 are denoted by the thick magenta line (above) and the thick blue line (below). (B) A portion of the spEM map (contoured at 2.5 σ) is shown (transparent grey surface) with the superimposed Taq core RNAP structure (left, with βi12 colored blue) and the fitted Eco RNAP model (right, with βi11 colored magenta). The view corresponds roughly to the reference view of the Eco RNAP model (top view), shown as a backbone worm and color-coded as follows: αI, αII, ω, gray; β', light pink, except βi6 is red; β, light cyan, except βi4 is green, βi9 is orange, and βi11 is magenta.

doi:10.1371/journal.pbio.1000483.g007
with the DNA. The surface is already overall acidic (Figure 5, front view), so it seems unlikely that phosphorylation at this site affects RNAP function by affecting interactions with the downstream DNA.

Conclusions

An understanding of the basic principles of transcription and its regulation has been garnered largely through detailed study of the transcription system of one organism, *Eco*, which has served as a model for understanding transcription at the molecular and cellular level for more than four decades. The detailed and comprehensive structural description of *Eco* core RNAP and an *Eco* RNAP TEC presented here sheds new light on the interpretation of previous biochemical and genetic data. Moreover, the molecular models provide a structural framework for designing future experiments to investigate the function of the *Eco* RNAP lineage-specific insertions and their role in the *Eco* transcription program, allowing a fuller exploitation of *Eco* as a model transcription system.

Materials and Methods

Crystallization and Structure Determination of *Eco* RNAP β2-βi4

*Eco* β2-βi4 was amplified by the polymerase chain reaction from the *Eco* spoB expression plasmid pRL706 [49] and cloned between the NdeI and BamHI sites of a pET28a-based expression plasmid, creating pSKB2(10-His)Ecoβ2-βi4, encoding *Eco* β2-βi4 with an N-terminal PreScission protease (GE Healthcare) cleavable His10-tag. The pSKB2(10-His)Ecoβ2-βi4 was transformed into *Eco* BL21 (DE3) cells. After growing transformed cells in LB medium with kanamycin (50 μg/ml) at 37 °C to an A600 = 0.6, isopropyl β-D-1-thiogalactopyranoside was added to a final concentration of 1 mM and cells were grown for an additional 3 h at 37 °C. Cells were harvested by centrifugation, resuspended in lysis buffer (20 mM Tris-HCl, 0.5 M NaCl, 0.5 mM β-mercaptoethanol, 5% v/v glycerol, 0.5 mM phenylmethanesulphonylfluoride), lysed in a continuous-flow French press (Avestin), and clarified by centrifugation. The protein was purified by HiTrap Ni2⁺-chelating affinity chromatography (GE Healthcare) and the His10-tag was removed using PreScission protease (GE Healthcare). The sample was further purified by a second, subtractive HiTrap Ni2⁺-chelating affinity chromatography step to remove uncleaved His10-tagged protein and the His10-tag released from the cleaved product, and gel filtration chromatography (Superdex 75, GE Healthcare). The purified protein was concentrated to 17 mg/ml by centrifugal filtration (VivaScience) and exchanged into storage buffer (10 mM Tris-HCl, pH 8.0, 0.15 M NaCl, 1 mM DTT), and stored at –80 °C. Selenomethionyl-substituted protein was prepared by suppression of methionine biosynthesis [50] and purified by using similar procedures. Reductive methylation of lysine residues was performed as described [20].

Crystals were grown at 22°C in sitting drops using vapor diffusion by mixing equal volumes of protein solution (0.5 μl at 6 mg/ml in storage buffer) and crystallization solution (0.2 M potassium-sodium tartrate, 20% PEG3350). Crystals (irregular plates) appeared after a few days and grew to a maximum size of about 200×100×50 μm in 1 wk. Crystals were prepared for cryocrystallography by a quick soak in cryo-solution (0.2 M potassium-sodium tartrate, 35% PEG3350), then flash frozen and stored in...
Cloning, expression, and purification of the
E. coli 6–14 subunit.

The 6–14 subunit was produced in the presence of 75 mg/ml ampicillin in 3 ml of
 LB media. After overnight growth at 37°C, the culture was lysed and the protein
 was purified on a Q-Sepharose column (Pharmacia). The final yield was 6 mg/ml
 in the final eluted fraction.

Cryoprotectant.

The protein was concentrated to 1 mg/ml and was then mixed with 100 mM
 potassium phosphate, 500 mM sucrose, EDTA, and 5 mM DTT to give a final
 concentration of 100 mg/ml. The solution was then divided into aliquots and
 frozen in liquid nitrogen.

Cryo-EM Reconstruction of Eco RNAP by Single-Particle
Averaging

The Eco RNAP was negatively stained with 4% uranyl acetate on a carbon film.

The crystals of Eco RNAP were streaked onto a 300-mesh carbon-coated Cu grid
 and stained with 4% uranyl acetate. The grid was then imaged at 120 kV at 1.5 M
 magnification and digitized using a Tecnai F20 transmission electron microscope.

The micrographs were digitized at a 14 Mpixel resolution and were analyzed
 using the SPARX software package [25] with a stop-band frequency of 0.28
 Å−1. For further analysis, the map was Fourier filtered using an aperiodic
 tangent pass filter [24] as implemented in the SPARX software package [25] with
 a stop-band frequency of 0.28 and a fall-off of 0.45.

Sequence Alignments

Alignments for the Eco lineage-specific insertions (see Datasets 1–3) were created
 using the bacterial lineage-specific insertions alignments from Lane et al. [3] as
 a starting point. The final alignments were created by iterative cycles in which
 sequences that did not match the Eco domains were removed, followed by
 realignment with MUSCLE [65] or PCMA [66].

Accession Numbers

Electron Microscopy Data Bank: The single-particle cryoEM reconstruction
 volume has been deposited under ID code EMD-5169. Protein Data Bank:
 Atomic coordinates and structure factors for Eco RNAP β2-β4 have been
 deposited under accession code 3L1T. The EM-fitted coordinate model of Eco
 core RNAP has been deposited under accession code 3L10. The coordinates of
 the Eco RNAP TEC model are available in the Supporting Information
 (Dataset S4).

Supporting Information

Dataset S1 beta-i9_blast_to_fas_to_aln_man4_cull.msf
  – Sequence alignment (msf format) containing 307 non-redundant β9
  sequences.
  Found at: doi:10.1371/journal.pbio.1000483.s001 (0.07 MB TDS)

Dataset S2 beta-i4_blast_to_fas_to_aln_man5_cull.msf
  – Sequence alignment (msf format) containing 316 non-redundant
  β4 sequences (only Eco-like β4 sequences comprising two BBM2
  domains).
  Found at: doi:10.1371/journal.pbio.1000483.s002 (0.07 MB TDS)

Dataset S3 beta-i11_blast_to_fas_to_aln_man4_cull.msf
  – Sequence alignment (msf format) containing 310 non-redundant
  β11 sequences.
  Found at: doi:10.1371/journal.pbio.1000483.s003 (0.07 MB TDS)
**Dataset S4 Eco_TEC_model.pdb – Coordinates (PDB format) of the Eco TEC model.**

Found at: doi:10.1371/journal.pbio.1000483.s004 (2.22 MB TXT)

**Figure S1 Eco β2-βi4 electron density map.** Stereo view of the 1.6 Å-resolution 2|Fo|–|Fc| map, contoured at 1.5 σ. The model is shown as sticks, with nitrogen atoms colored blue, oxygen atoms red, and carbon atoms colored according to Figure 2B. Water molecules are represented as red spheres. Shown is the region surrounding dimethylated [20] K324.

Found at: doi:10.1371/journal.pbio.1000483.s005 (2.07 MB TIF)

**Figure S2 Comparison of Tag β2 and Eco β2-βi4.** The two structures were superimposed over 100 χ-carbon positions (excluding flexible loops connecting secondary structural elements), yielding a root-mean-square-deviation of 1.68 Å. Other than the insertion of βi4 in Eco, significant differences in the β2 structures include: (i) the loop connecting the first two β-strands of the β2 domain, where Eco has a 5-residue insertion (Eco β residues 164–168, disordered in the structure), and (ii) the loop connecting the last two α-helices of the β2 domain, which includes a 7-residue insertion present in Tag β (Tag β residues 293–299; Figure 2A).

Found at: doi:10.1371/journal.pbio.1000483.s006 (5.47 MB TIF)

**Figure S3 Eco β⅃ap-β⅃9 electron density map.** Stereo view of the 3.0 Å-resolution 2|Fo|–|Fc| map, contoured at 1.0 σ. The model is shown as sticks, with nitrogen atoms colored blue, oxygen atoms red, and carbon atoms colored according to Figure 3B. Shown is a region of the β⅃9 ladder helices.

Found at: doi:10.1371/journal.pbio.1000483.s007 (2.90 MB TIF)

**Figure S4 Image analysis.** (A) Unprocessed electron micrograph of a field of Eco RNAP molecules preserved in vitreous ice. Selected particles are circled. (B) Distribution of image orientations, plotted as a polar-angle diagram, viewed along the 0 = 0° axis. (C) Fourier shell correlation [67,68] as a function of spatial frequency.

Found at: doi:10.1371/journal.pbio.1000483.s008 (1.54 MB TIF)

**Figure S5 Back, bottom, channel, and front views of spEM density and fit of Eco RNAP model.** For each view, the left image shows the spEM density map (grey surface, contoured at 2.5 σ), and the right image shows the spEM density map (grey transparent surface) with the fitted Eco RNAP homology model superimposed (excluding θ, the C-terminal 41 residues of β’ and β9). The Eco RNAP homology model is shown as a backbone worm, color-coded as in Figure 4.

Found at: doi:10.1371/journal.pbio.1000483.s009 (7.72 MB TIF)

**Figure S6 β-side, bottom, β-side, and top views of spEM density and fit of Eco RNAP model.** For each view, the left image shows the spEM density map (grey surface, contoured at 2.5 σ), and the right image shows the spEM density map (grey transparent surface) with the fitted Eco RNAP homology model superimposed (excluding θ, the C-terminal 41 residues of β’ and β9). The Eco RNAP homology model is shown as a backbone worm, color-coded as in Figure 4.

Found at: doi:10.1371/journal.pbio.1000483.s010 (8.62 MB TIF)

**Figure S7 Structural features of Eco β9.** Two views of Eco β9 are shown: The left column shows the “front” view (the side facing the “hook”), and the right column shows the “back” view (the side away from the “hook”). The top row shows the backbone ribbon. The middle row shows the structure (with transparent molecular surface) colored in a gradient according to the Blosum 62 information score (as determined by the program PFAAT [70]) calculated from an alignment of 307 non-redundant β9 sequences (see Supporting Information). The color gradient covers scores from 0 to 1 (0, white; 0.5, yellow; 1.0, red). Individual residues with score ≥0.75 are labeled. Underlined residues denote residues with significant solvent accessibility. The bottom row shows the molecular surface colored according to the electrostatic surface distribution of the solvent-accessible surface in units of KT (−5, red; 0, white; +5, blue), as calculated by APBS [69].

Found at: doi:10.1371/journal.pbio.1000483.s011 (6.13 MB TIF)

**Figure S8 Details of ab initio-predicted Eco β11 structure.** (A) Sequence context of Eco RNAP β11. The secondary structure for the predicted Eco β11 structure (determined using the Robetta server (http://robetta.bakerlab.org/)) is indicated directly below the sequence (filled rectangles denote α-helices). Above the number scale, black lines denote the sequence regions common to all bacterial RNAPs [3]. Gaps in the β11 sequence with numbers above denote the location and residue length of insertions in an alignment of 310 non-redundant β11 sequences (see Supporting Information). The insertions all occur in loops connecting the helices. The extent of Eco β11 is denoted by the thick magenta line (above). (B) Backbone ribbon of the predicted Eco β11 structure. The grey spheres mark α-carbon positions surrounding the insertions from the sequence alignment. The numbers pointing to each insertion point denote the insertion length. (C) The predicted Eco β11 structure is colored in a gradient according to the Blosum 62 information score (as determined by the program PFAAT [70]) calculated from the alignment of 310 non-redundant β11 sequences (see Supporting Information). The color gradient covers scores from 0 to 1 (0, white; 0.5, yellow; 1.0, red). Individual residues with score ≥0.75 are labeled. Nearly all of the conserved hydrophobic residues are buried in the hydrophobic core of the structure. Two solvent-accessible polar residues (R1142 and D1166) form an apparently conserved salt-bridge that may stabilize the structure.

Found at: doi:10.1371/journal.pbio.1000483.s012 (3.18 MB TIF)

**Table S1 Crystallographic statistics for Eco RNAP β⅃ap-β⅃9 crystals.**

Found at: doi:10.1371/journal.pbio.1000483.s013 (0.04 MB DOC)

**Acknowledgments**

We thank Deena Oren of The Rockefeller University Structural Biology Resource Center (RU-SBRC) for expert assistance. We thank Wuxian Shi at the National Synchrotron Light Source beamline X3A for assistance with data collection. We thank K.D. Derr, Bill Rice, and Ruben Diaz-Avalos from the New York Structural Biology Center (NYSBC) for helpful technical support in using the Tecnai F20 electron microscope.

**Author Contributions**

The author(s) have made the following declarations about their contributions: Conceived and designed the experiments: NJO KAFT RL FJA SAD. Performed the experiments: NJO JB KAFT RL FJA SAD. Contributed reagents/materials/analysis tools: FJA. Wrote the paper: FJA SAD. Technical support in using the Tecnai F20 electron microscope: Technical support in using the Tecnai F20 electron microscope.

**References**

1. Jokerst RS, Weeks JR, Zehring WA, Greenleaf AL (1989) Analysis of the gene encoding the largest subunit of RNA polymerase II in Deosphilale. Mol Gen Genet 215: 266–275.

2. Sweetser D, Nonet M, Young RA (1987) Prokaryotic and eukaryotic RNA polymerases have homologous core subunits. Proc Natl Acad Sci U S A 84: 1192–1196.
17. Severinov K, Kashlev M, Severinova E, Bass I, McWilliams K, et al. (1994) A
16. Kim DE, Chivian D, Baker D (2004) Protein structure prediction and analysis
11. Gross CA, Chan CL, Lonetto MA (1996) A structure/function analysis of
10. Chlenov M, Masuda S, Murakami KS, Nikiforov V, Darst SA, et al. (2005)
20. Rayment I (1997) Reductive alkylation of lysine residues to alter crystallization
18. Artsimovitch I, Svetlov V, Murakami KS, Landick R (2003) Co-overexpression
29. Zhang G, Darst SA (1998) Structure of the Escherichia coli RNA polymerase
27. Tan RK-Z, Devkota B, Harvey SC (2008) YUP.SCX: coaxing atomic models
26. Kastner B, Fischer N, Golas MM, Sander B, Dube P, et al. (2008) GraFix:
24. Basokur AT (1998) Digital filter design using the hyperbolic tangent functions.
23. Opalka N, Mooney RA, Richter C, Severinov K, Landick R, et al. (1999) Direct
20. Rayment I (1997) Reductive alkylation of lysine residues to alter crystallization
12. Zhang G, Campbell EA, Minakhin L, Richter C, Severinov K, et al. (1999)
10. Chlenov M, Masuda S, Murakami KS, Nikiforov V, Darst SA, et al. (2005)

60. Opalka N, Chlenov M, Chacon P, Rice WJ, Wriggers W, et al. (2003) Structure and function of the transcription elongation factor GreB bound to bacterial RNA polymerase. Cell 114: 335–345.

61. Ludtke SJ, Baldwin PR, Chiu W (1999) EMAN: semi-automated software for high resolution single particle reconstructions. J Structural Biol 128: 82–97.

62. Craighead JL, Chang WH, Asturias FF (2002) Structure of yeast RNA polymerase II in solution: implications for enzyme regulation and interaction with promoter DNA. Structure (Camb) 10: 1117–1125.

63. Frank J, Radhakrishnan M, Penczek P, Zhu J, Li Y, et al. (1996) SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. J Struct Biol 116: 190–199.

64. Penczek P, Grassucci RA, Frank J (1994) The ribosome at improved resolution: new techniques for merging and orientation refinement in 3D cryo-electron microscopy of biological particles. Ultramicroscopy 53: 251–270.

65. Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5: 113.

66. Pei J, Sadreyev R, Grishin NV (2003) PCMA: fast and accurate multiple sequence alignment based on profile consistency. Bioinformatics 19: 427–428.

67. Saxton WO, Baumeister W (1982) The correlation averaging of a regularly arranged bacterial-cell envelope protein. J Microsc 127: 127–138.

68. Harauz G, van Heel M (1986) Exact filters for general geometry three-dimensional reconstruction. Optik 73: 146–156.

69. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci U S A 98: 10037–10041.

70. Caffrey DR, Dana PH, Mathur V, Ocano M, Hong EJ, et al. (2007) PFAAT version 2.0: a tool for editing, annotating, and analyzing multiple sequence alignments. BMC Bioinformatics 8: 381.