Three-dimensional EM structure of an intact activator-dependent transcription initiation complex

Brian P. Hudson^{a,b}, Joel Quispe^c, Samuel Lara-González^a, Younggyu Kim^{a,d,e}, Helen M. Berman^a, Eddy Arnold^{a,b}, Richard H. Ebright^{a,d,e}, and Catherine L. Lawson^{a,1}

^aDepartment of Chemistry and Chemical Biology and ^bCenter for Advanced Biotechnology and Medicine, Rutgers University, 610 Taylor Road, Piscataway, NJ 08854; ^cNational Resource for Automated Molecular Microscopy, The Scripps Research Institute, CB 129, 10550 North Torrey Pines Rd, La Jolla, CA 92037; and ^dWaksman Institute and ^eHoward Hughes Medical Institute, Rutgers University, Piscataway, NJ 08854

Edited by Seth A. Darst, Rockefeller University, New York, NY, and approved October 05, 2009 (received for review August 03, 2009)

We present the experimentally determined 3D structure of an intact activator-dependent transcription initiation complex comprising the *Escherichia coli* catabolite activator protein (CAP), RNA polymerase holoenzyme (RNAP), and a DNA fragment containing positions -78 to $+20$ of a Class I CAP-dependent promoter with a CAP site at position -61.5 and a premelted transcription bubble. A 20-Å electron microscopy reconstruction was obtained by iterative projection-based matching of single particles visualized in carbon-sandwich negative stain and was fitted using atomic coordinate sets for CAP, RNAP, and DNA. The structure defines the organization of a Class I CAP-RNAP-promoter complex and supports previously proposed interactions of CAP with RNAP α subunit C-terminal domain (α CTD), interactions of α CTD with σ^{70} region 4, interactions of CAP and RNAP with promoter DNA, and phased-DNA-bend-dependent partial wrapping of DNA around the complex. The structure also reveals the positions and shapes of species-specific domains within the RNAP β' , β , and σ^{70} subunits.

catabolite activator protein | deoxyribonucleic acid |
RNA polymerase holoenzyme | gene regulation | electron microscopy

The *Escherichia coli* catabolite activator protein (CAP) is a classical model for structural and mechanistic studies of transcription activation (1, 2). CAP is a global regulator of gene expression in *E. coli* (1–3) and is the most extensively studied member of a large family of structurally related transcriptional regulators that mediate response to a broad spectrum of environmental signals in bacteria (1–4).

CAP is a dimer of two 23-kDa subunits. Upon binding to the allosteric effector cAMP, CAP binds to a 22-bp inverted-repeat DNA site and activates transcription at adjacent or overlapping promoters (1, 2, 5, 6). Structures of CAP-DNA complexes indicate that CAP bends its DNA site by $\approx 80^\circ$ (5, 6).

Transcription activation at the simplest CAP-dependent promoters requires only three macromolecular components—CAP, RNA polymerase holoenzyme (RNAP; subunit composition $\alpha^1\alpha^1\beta\beta'\omega\sigma^{70}$), and promoter DNA—and requires only one DNA site for CAP (1, 2). At “Class I” CAP-dependent promoters, typified by the *lac* promoter, CAP interacts with a DNA site located upstream of the core promoter and makes a set of protein-protein interactions with RNAP that facilitate association of RNAP with the promoter to form an RNAP-promoter closed complex (1, 2). At “Class II” CAP-dependent promoters, typified by the *gal* promoter, CAP interacts with a DNA site overlapping the core promoter and makes a different set of protein-protein interactions with RNAP that facilitate isomerization of the RNAP-promoter closed complex to the catalytically competent RNAP-promoter open complex (1, 2).

At Class I CAP-dependent promoters, CAP activates transcription by making protein-protein interactions with the RNAP α subunit C-terminal domain (α CTD), a 9-kDa independently folded domain that is flexibly tethered to the remainder of RNAP (1, 2, 7, 8). Interaction of CAP with α CTD facilitates binding of α CTD—and, through it, the remainder of RNAP—to promoter DNA, and

thereby stimulates transcription initiation. The interaction between CAP and α CTD is mediated by residues 156–164, 209 of the downstream subunit of the CAP dimer (“activating region 1;” AR1) and residues 285–288, 315, 317 of α CTD (“287 determinant”) (1, 2, 8). The interaction between α CTD and DNA is mediated by residues 265, 294–302 of α CTD (“265 determinant”) and the DNA minor groove (1, 2, 8). The interaction between α CTD and DNA can occur with any DNA sequence, but is most favorable with A/T-rich DNA sequences (1, 8, 9).

At Class I CAP-dependent promoters where the DNA site for CAP is centered at position -61.5 , such as the *lac* promoter, the interaction between CAP and α CTD places α CTD adjacent to the RNAP σ^{70} subunit and permits functional protein-protein interaction between α CTD and σ^{70} (2, 10). The interaction between α CTD and σ^{70} is mediated by residues 257–261 of α CTD (“261 determinant”) and residues 573–604 within the module of σ^{70} responsible for recognition of the promoter -35 element, σ^{70} region 4 (σ R4; “596 determinant”) (2, 10, 11).

RNAP contains two copies of α CTD. At Class I CAP-dependent promoters, one α CTD protomer—interchangeably α CTD^I or α CTD^{II}—interacts with CAP (1, 2, 12–15). The other α CTD protomer interacts nonspecifically with upstream DNA (1, 2, 12, 13, 16).

The crystal structure of a subassembly comprising CAP, α CTD, and DNA provided a high-resolution structural description of the interaction between a transcriptional activator and its functional target within the general transcriptional machinery and confirmed predicted protein-protein and protein-DNA interactions (8). The crystal structure, in conjunction with additional structural, biochemical, and genetic data, also enabled construction of a structural model for the intact Class I CAP-dependent transcription activation complex (2). However, the accuracy of the model could not be independently assessed because experimental data regarding the overall shape of the complex were not available.

Although CAP-RNAP-promoter complexes are easily assembled in solution (12, 17), they have proven refractory to crystallization efforts, likely owing to large size and to conformational heterogeneity. We therefore sought to define structures of these ≈ 500 -kDa complexes using single particle electron microscopy (EM) reconstruction, an increasingly important tool in

Author contributions: B.P.H., H.M.B., E.A., R.H.E., and C.L.L. designed research; B.P.H., J.Q., and C.L.L. performed research; S.L., Y.K., and R.H.E. contributed new reagents/analytic tools; B.P.H. and C.L.L. analyzed data; and B.P.H., R.H.E., and C.L.L. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The EM reconstruction volume has been deposited in the Electron Microscopy Data Bank, www.emdatabank.org (EMD ID code EMD-5127); the EM fitted coordinate model has been deposited in the Protein Data Bank, www.pdb.org (PDB ID code 3iyd).

¹To whom correspondence should be addressed. E-mail: cathy.lawson@rutgers.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0908782106/DCSupplemental.

Fig. 1. Promoter DNA construct. The consensus DNA site for CAP, the consensus DNA site for α CTD, and core promoter elements are labeled and highlighted in gray. The bases altered to create noncomplementarity in positions -11 to $+2$, resulting in an artificial transcription bubble, are shown in lowercase.

structural biology research (18). This method is particularly informative in instances when high-resolution crystal structures are available to aid interpretation of low-resolution envelopes. Here, we present the EM structure of an *E. coli* Class I CAP-dependent transcription activation complex determined at 20 Å resolution, its interpretation using high-resolution X-ray crystal structures, and a comparison to the published model.

Results

Design, Assembly, and Imaging of a Class I CAP-RNAP-Promoter Complex. The promoter DNA construct was designed based on positions -78 to $+20$ of the Class I CAP-dependent promoter *lac* (Fig. 1). The DNA sites for CAP and α CTD were replaced by consensus DNA sites for CAP and α CTD (AAATGTGATCTAGATCACATTT and AAAAAA), as in (8). The -10 element was replaced by a consensus -10 element (TATAAT), and the discriminator element was replaced by a consensus-type discriminator element (CGC) (19). Positions -11 to $+2$ were made noncomplementary to create an artificial transcription bubble (Fig. 1).

Incubation of CAP and RNAP with the promoter DNA construct in the presence of cAMP yielded CAP-RNAP-promoter complexes (Fig. S1), and unbound RNAP was removed using heparin-Sepharose. No further purification was required, since excess unbound CAP (47 kDa) and multiRNAP-containing aggregates ($>1,000$ kDa) are readily distinguished from CAP-RNAP-promoter complexes (500 kDa) in EM images. Complexes were visualized in a carbon-sandwich uranyl formate negative-stain preparation, as described by Ohi et al. (20). Even though DNA has weaker contrast than protein in negative stain (21–24), this preparation method has yielded the most structural detail thus far of the complex. Three-dimensional (3D) reconstructions from negative stain typically yield accurate (to within a few Å) localization of protein domain centers relative to each other and permit mapping of molecular interaction sites at a 10–20 Å resolution level (25).

Reconstruction and Coordinate Model Fitting. The 3D reconstruction was obtained using projection matching, with a strongly Gaussian blurred representation of *Thermus thermophilus* RNAP holoenzyme (26) as the starting volume (Fig. S2A and B). *T. thermophilus* RNAP and *E. coli* RNAP share $\approx 40\%$ sequence identity and have roughly the same shape and size, but also have well-characterized species-specific differences (27–31) (Fig. S2C). Trial 3D reconstructions were judged in part by the appearance of expected *E. coli* species-specific domains—in particular, β dispensable region 1 (β DR1; also known as SI1) (27–29), β dispensable region 2 (β DR2; also known as SI2) (27–29, 31), β' trigger-loop nonconserved domain (β' GNCD; also known as SI3) (28–30), and the σ^{70} nonconserved-region insert within σ^{70} region 2 (σ^{70} NCR) (32). The final reconstruction was created from 13,841 particles in 280 classes (Fig. S3). The map resolution, estimated by Fourier Shell Correlation, is 20 Å (Fig. S3A).

To aid interpretation of the map volume, a coordinate model of the complex was generated in four steps. First, the map was fitted with coordinates from crystal structures of CAP- α CTD-DNA (8), α CTD- σ R4-DNA (see SI Text), *T. thermophilus* RNAP holoenzyme (26), *E. coli* α subunit N-terminal domain dimer (α NTD₂) (33), *E. coli* σ^{70} regions 1.2–2 (32), and *E. coli* β' GNCD (30) using UCSF Chimera (34). Fitting statistics are provided in Table S1. Coordinates for σ^{70} region 1.1 (35) were not fitted to the map, because density could not be unambiguously assigned for this small domain. Second, a homology model for *E. coli* RNAP was generated from the heterogeneous fitted RNAP fragments. Third, a complete model for the DNA was generated based on the fitted crystallographic structures and fitted DNA from the published Class I CAP-RNAP-promoter model (2). Fourth, the coordinate model was refined against the experimental map using a simulated annealing protocol that primarily shifts domain-sized fragments divided along natural boundaries as rigid bodies (36). Refinement improved the model-map correlation coefficient from 0.77 to 0.85 with root-mean-square shift of 5.5 Å for all atoms.

Structure of the Class I CAP-RNAP-Promoter Complex. The EM reconstruction defines the shapes and relative positions of the three components of the complex: CAP, RNAP, and promoter DNA (Fig. 2). When the map is contoured to match the expected volume of the ternary complex, one observes substantially complete density for RNAP and partial densities at expected positions of the downstream subunit of CAP and of one promoter of α CTD (Fig. 2A and C, solid purple). When the map is contoured at a lower level, one observes additional density at expected positions of the species-specific RNAP inserts β DR2 and β' GNCD, essentially complete density at the expected position and with the expected boxy shape of the CAP dimer, essentially complete density at the expected position and with the expected ellipsoid shape of α CTD, and partial density at expected positions and with the expected extended shapes of promoter DNA segments upstream and downstream of the transcription bubble (Fig. 2A–D, semitransparent purple). With further reduced contour level, additional density is observed at expected positions and with the expected extended shapes for promoter DNA [Fig. 2A and C, gray mesh; compare with DNA (red) in Fig. 2B and D]. Weaker protein density is interpreted to arise from positional and/or occupancy variability of the affected regions within the particle population; weaker DNA density is expected, owing to reduced contrast (21–24).

The EM reconstruction is similar to our published model of the Class I CAP-RNAP-promoter complex (2) and supports all major features of the published model, including interactions of CAP with RNAP α CTD mediated by AR1 of CAP and the 287 determinant of α CTD, interactions of α CTD with σ R4 mediated by the 261 determinant of α CTD and the 596 determinant of σ R4, and interactions of CAP and RNAP with promoter DNA and placement of positions -40 to -100 of the upstream promoter region in close proximity to RNAP (Figs. 2 and 3, and Fig. S4). Relative to RNAP, the position of CAP in the EM reconstruction differs from the position of CAP in the published model by a small translation (6 Å) and a small rotation (23°) (Fig. S4). The published model was constructed based on the crystal structure of the CAP- α CTD-DNA complex (8) and the crystal structure of an RNAP-DNA complex (37), using restraints derived from genetic and biochemical experiments defining protein-protein contacts (2). The close correspondence of the experimentally obtained EM structure and the published model validates the approach of modeling large assemblies by combining crystal structures of subassembly fragments and experimentally derived restraints.

In the EM structure, the α CTD protomer that contacts CAP is approximately equidistant from the C terminus of α NTD^I and the C terminus of α NTD^{II} (≈ 50 Å, in each case), and in principle

Fig. 2. Class I CAP-RNAP-promoter complex EM reconstruction and fitted model (stereo pairs). (A) EM map density shown contoured at three levels (5σ , solid purple; 2.6σ , semitransparent purple; 1.0σ , gray mesh). (B) Ribbon representation of the fitted model in same view orientation, overlaid with the 2.6σ map contour (semitransparent gray). DNA and cAMP bound to CAP are in red; RNAP subunits are colored as follows: α^I , yellow; α^{II} , green; β , cyan; β' , pink; ω , gray; and σ^{70} , orange. View orientations in C and D are rotated by $\approx 180^\circ$ about the y axis relative to panels A and B.

could be connected to either αNTD^I or αNTD^{II} (Fig. 3), consistent with the observation that both αCTD^I and αCTD^{II} can support CAP-dependent transcription at *lac* (1, 2, 12, 13). In Figs. 2 and 3, the αCTD protomer that contacts CAP has been connected to αNTD^I based on protein-protein cross-linking data indicating that the αCTD protomer connected to αNTD^I preferentially, by a factor of $\approx 3:1$, makes contact with CAP. Biochemical data indicate that the αCTD protomer that does not make contact with CAP interacts nonspecifically with the DNA minor groove at positions -73 , -83 , and -93 (2, 12) and is available, in principle, to interact with a second activator (1, 2, 13, 16). Although there is no density feature in the EM map that is clearly assignable to the second αCTD protomer, there is sufficient space in the EM reconstruction to accommodate a second αCTD protomer near position -73 in the DNA minor groove. The promoter DNA segment upstream of the DNA site for CAP is ≈ 10 Å closer to the RNAP αNTD_2 dimer in the EM

Fig. 3. Class I CAP-RNAP-promoter complex EM reconstruction and fitted model: interactions between CAP, αCTD , and σR4 (stereo pair; density contoured at 2.6σ). Coordinates for CAP, αCTD , σR4 , and DNA in this region are based on crystal structures of CAP- αCTD -DNA (8) and αCTD - σR4 -DNA. CAP activating region 1 (AR1) is shown in yellow; the 287, 265, and 261 determinants of αCTD are shown in green, red, and blue, respectively; and the 596 determinant of σR4 is shown in pink. Ribbon colors are as in Fig. 2. The protomer of αCTD adjacent to CAP has been linked to αNTD^I , but it alternately could have been linked to αNTD^{II} (alternate linkage position indicated by black sphere at lower right).

reconstruction than in the published model (Fig. S4); the closest approach is ≈ 16 Å between DNA position -73 and αNTD^{II} residues 14–21, 210–215.

In both the EM reconstruction and the published model, small deformations of the DNA were introduced to either side of the -35 element, leading to bending of the helical path by $\approx 15^\circ$. In the EM structure, DNA bending was required to satisfy the observed placement of CAP relative to RNAP. In the published model, DNA bending was required to satisfy noninterpenetration and proximity constraints. The EM structure confirms the prediction of the published model that, owing to consecutive phased bends in the DNA, the binding of CAP at position -61.5 places the entire upstream promoter region, from positions -100 to -40 , in close proximity to RNAP and results in partial wrapping of DNA around the complex (Figs. 2–4). The phased DNA-bend-dependent partial wrapping yields an apparent over-

Fig. 4. Class I CAP-RNAP-promoter complex EM reconstruction and fitted model: inferred path of DNA in 2D projection, apparent overall DNA compaction, and apparent overall DNA bend angle (map density contoured at 1.0σ). The apparent compaction, ≈ 19 nm, is the difference between (i) the measured distance (d) between the points where DNA enters and exits the complex and (ii) the calculated length of the DNA segment ($98\text{-bp} \times 3.4$ Å rise/bp); the apparent bend angle is defined as the deviation of the DNA helix axis from a straight line (parameters defined as in ref. 55). The fitted coordinate model DNA (gray) was extended on each end by addition of 50-bp of canonical B-form DNA (black) generated using 3DNA (50).

Fig. 5. Class I CAP-RNAP-promoter complex EM reconstruction and fitted model: *E. coli* species-specific RNAP domains. (A) β' trigger-loop nonconserved domain (β' GNCD) with β' GNCD SBHMa domain in orange; β' GNCD SBHMB domain in yellow; β' trigger loop in green; β' jaw in blue (stereo pair). (B) β DR1. (C) β DR2. (D) β residues 1126–1179 (as well as β N-terminal residues 1–9). (E) σ^{70} NCR with σ^{70} NCR insert in green and 20-residue negatively charged loop that is disordered in the crystal structure of the σ^{70} fragment (32) indicated as a black dashed line. In each panel, insertion starting and ending points are indicated by black spheres, the EM map is contoured at 2.6σ , and other ribbon colors are as in Fig. 2. Stereo pairs for B–E are shown in Figs. S6 and S7.

all DNA compaction of ≈ 19 nm (equivalent to ≈ 55 -bp) and apparent overall DNA bend angle of $\approx 130^\circ$ (Fig. 4).

Species-Specific RNAP Domains. The EM map reveals the locations of *E. coli* species-specific insertions within β' , β , and σ^{70} , and in some cases permits modeling of orientations and potential interactions of inserted domains (Fig. 5 and Figs. S5–S7). The *E. coli* species-specific insertions within β' , β , and σ^{70} are unequivocally identifiable as large density features that are not fitted by coordinate models of CAP, α CTD, and *T. thermophilus* RNAP and that are at expected sequence positions relative to the fitted coordinate model of *T. thermophilus* RNAP.

The *E. coli* RNAP β' subunit trigger loop (TL), alternatively known as the G-loop, contains a 187-residue nonconserved domain (β' GNCD; residues 943–1130) (Fig. S2C). The TL is a flexible structural element of the RNAP active center that undergoes transitions between a closed (or folded) conformation and an open (or unfolded) conformation (38, 39). It has been proposed that the TL cycles between open and closed conformations during each nucleotide-addition step and that TL conformational cycling is important for discrimination between complementary and noncomplementary ribonucleotide substrates, for discrimination between ribonucleotides and deoxyribonucleotides, and, possibly, for catalysis (38–40). A crystal structure of *E. coli* β' GNCD indicates that β' GNCD is rod-shaped and comprises two tandem repeats of a sandwich-barrel hybrid motif: SBHMa and SBHMB (30). The crystal structure could be fitted to the EM map in an orientation that (i) placed the β' GNCD N and C termini to connect to the TL open

conformation (coordinates as in PDB ID 1zyr) (41), (ii) placed SBHMa into density near the RNAP jaw (β' residues 1151–1216) (42), and (iii) placed SBHMB into density near the RNAP jaw and near β DR1 (β residues 223–339) (Fig. S5A). In the resulting fitted coordinate model, residues 977–979 and 996–999 of SBHMa and residues 1025–1030, 1093–1104, and 1124–1121 of SBHMB are positioned potentially to make direct contact with the RNAP jaw (Fig. S6B). In the resulting fitted model, SBHMB is positioned ≈ 20 Å from the DNA duplex downstream of the transcription bubble (Fig. 5A and Fig. S5B). The fitted placement of β' GNCD is similar to that in the schematic model proposed in (30), but differs in that SBHMB is not placed directly adjacent to DNA. In a parallel trial, we attempted to connect the fitted β' GNCD crystal structure to the TL closed conformation (coordinates as in PDB ID 2o5j) (39), but this resulted in strained conformations and in unraveling of the TL helices during model refinement. We conclude that the fitted position of β' GNCD is likely to be compatible with the TL open conformation but is likely to be incompatible with the TL closed conformation. We infer that β' GNCD is likely to cycle between at least two alternative positions—one characteristic of the TL open conformation and one characteristic of the TL closed conformation—in concert with opening and closing of the TL, and we speculate that β' GNCD cycling may help couple opening and closing of the TL to RNAP jaw function, RNAP β DR1 function, and/or RNAP-DNA interactions.

The *E. coli* RNAP β subunit contains a 117-residue dispensable region (β DR1; residues 223–339), a 99-residue dispensable region (β DR2; residues 938–1036), and one smaller nonconserved region

(β residues 1126–1179) (Fig. S2C). High-resolution structures are not available for these regions. Density assignments for these regions are indicated with cyan spheres in Fig. 2B and D, and more detailed views are provided in Fig. 5B–D. The map densities for β DR1 and β residues 1126–1179 are as strong as densities for RNAP conserved regions. For β DR2, a volume consistent with the expected mass is obtained only at lower contour levels, suggesting partial disorder. By comparison, in an EM structure of *E. coli* RNAP core (27), β DR1 is partially disordered, β DR2 is ordered, and β 1126–1179 appears to be disordered.

Compared to *T. thermophilus* σ^A , the *E. coli* σ^{70} subunit has a 214-residue nonconserved region between σ region 1.2 and σ region 2 (σ NCR; residues 138–351) (Fig. S2C). A crystal structure of a σ^{70} fragment containing σ region 1.2, σ^{70} σ NCR, and σ region 2 (32) exhibits a pronounced “hook” shape that matches and can be fitted to a density feature in the expected location of the EM map (Fig. S7A). A highly negatively charged 20-residue loop within σ^{70} NCR (residues 192–211) that is disordered in the σ^{70} fragment crystal structure (32) also is not visible in the EM structure, presumably due to disorder (Fig. 5E). The EM structure is consistent with a previous suggestion (32) that the negatively charged loop may be near to promoter DNA, particularly the promoter –17 to –10 region, and, as such, potentially may influence promoter recognition and/or promoter melting (Fig. 5E and Fig. S7B).

Discussion

The 20-Å structure of an intact, complete *E. coli* Class I CAP-dependent transcription activation complex, assembled from CAP, RNAP, and promoter DNA with a designed artificial transcription bubble, was determined by calculating a 3D volume derived from thousands of 2D EM images of single particles. A coordinate model for the complex was generated by fitting crystal structures of subassemblies and components into the map, followed by homology modeling and refinement. The map plus its model interpretation define the overall shape and relative positions of CAP, RNAP, and DNA. The structure is consistent with the results of prior biophysical, biochemical, and genetic experiments and confirms conclusions of prior modeling based on the biophysical, biochemical, and genetic data.

This study establishes suitability of EM single-particle reconstruction for determining the overall shapes of multicomponent bacterial transcription complexes. The work described here represents the successful application of the method to an intact, complete activator-dependent transcription initiation complex bound to promoter DNA. Single-particle EM has also recently been applied to structure determination of *E. coli* σ^{54} RNAP in complex with an AAA+ activator (43) and *Bacillus subtilis* RNAP core in complex with elongation factor NusA (44). The methods defined in this study should be applicable to analysis of additional bacterial transcription complexes, to define binding sites for regulators of bacterial transcription, and to define the structural consequences of binding of regulators of bacterial transcription.

Methods

Samples. Full material preparation details are presented in the SI Text. RNAP (11.5 μ M) and DNA (10.9 μ M) were incubated for 15 min at 37 °C in 25 mM HEPES-NaOH (pH 8.0), 100 mM KCl, 10 mM MgCl₂, 1 mM 1,4-dithio-DL-threitol (DTT), and 200 μ M cAMP. After removal of unbound RNAP by treatment with 10 mg/mL heparin-Sepharose CL-6B (Amersham) for 15 min at 37 °C and filtration (Millex-GV, 0.22 μ m; Millipore), CAP (21.8 μ M) was added, and samples were incubated 15 min at 37 °C.

Electron Microscopy. Negatively stained EM specimens were prepared using a carbon-sandwich technique and uranyl-formate stain (20) (see SI Text for details). Images ($n = 349$; nominal magnification, $\times 50,000$; randomized defocus range, –0.5 to –1.5 mm) were acquired at the National Resource for Automated Molecular Microscopy (NRAMM), using a Tecnai F20 transmission EM at 120 keV

equipped with a Tietz F415 4 $\times$ 4k pixel CCD camera (15 μ m pixel) and a standard side entry room temperature stage, and were recorded using the Legikon automated EM package (45).

Single-Particle Reconstruction. Initial processing was performed using the NRAMM Appion pipeline (46). Particles were picked automatically using DoG (46), and contrast-transfer-function parameters were calculated using ACE (47). A phase-corrected stack of 32,816 particles was created with 3-fold pixel binning (80 $\times$ 80 pixels, 4.64 Å/pixel). The starting volume for reconstruction was prepared with EMAN pdb2mrc (48) using coordinates for *T. thermophilus* RNAP (PDBid 3dxj) (41), with Gaussian blurring to 60 Å resolution (Fig. S2B). The 3D reconstruction was generated in three stages: (i) the Appion-generated stack was filtered to exclude particles substantially different from the starting volume—e.g., particles comprising aggregates—via 12 cycles of EMAN projection-based matching particle classification (48) interleaved with SPIDER (49) correspondence analysis (procedure as described in (46)); SPIDER reference-free classification is performed on each EMAN class and the SPIDER class with highest correlation to the reference projection replaces the EMAN class, yielding a 13,490-particle reduced particle stack; (ii) 12 EMAN cycles were performed against the starting volume with the reduced particle stack; and (iii) 12 EMAN-SPIDER cycles were performed against the full particle stack. EMAN parameters in the final round were ang = 8 pad = 100 mask = 30 hard = 25 classkeep = 0.8 classiter = 8 phasecls. The Fourier shell correlation was determined using EMAN eotest and 0.5 cutoff. The final map was low pass filtered to 18 Å. The average map density was 0.00, with range –3.34 to +14.40. Map contour levels are expressed as σ values (raw values divided by the standard deviation, 0.90). The expected volume of the complex was calculated assuming 1.35 g/mL for 389-kDa RNAP and 47-kDa CAP and 1.7 g/mL for 64-kDa DNA.

Coordinate Model Fitting. Multiple high-resolution crystal structures were fitted to the EM map using UCSF Chimera (34). A composite model for CAP, α CTD, σ R4, and DNA (–78 to –28) was created by superposition of all common nonhydrogen atoms of two structures representing distinct subassemblies: CAP- α CTD-DNA (PDBid 1lb2) (8) and α CTD- σ R4-DNA (see SI Text). The composite model was placed into the map by fitting the CAP dimer into its assigned density. After *T. thermophilus* RNAP (PDBid 3dxj) (26) was fitted, the σ R4 domain was aligned to the composite model σ R4 position, and the trigger loop was replaced with open conformation coordinates (PDBid 1zyr) (41). Essentially equivalent placements were obtained for *E. coli* α NTD₂ (PDBid 1bdf) (33) and *E. coli* σ^{70} region 1.2–2 (PDBid 1sig) (32) by superimposition onto the fitted *T. thermophilus* RNAP or by map density fitting. *E. coli* β' GNCD (PDBid 2auk) (30) was manually fitted into density.

DNA coordinates for positions –23 to +20 were taken from the published model (2). DNA from the fitted composite model was extended downstream by 5 ideal B-DNA base pairs; tilt, roll, twist, and rise parameters for positions –36 to –23 were manually adjusted using 3DNA (50) to produce a smooth bend of $\approx 15^\circ$ and yield superimposed –23 bp positions for the two DNA model segments.

RNAP Homology Modeling. Sequence alignments from *E. coli*, *T. thermophilus*, and other organisms were created for each RNAP subunit using CLUSTALW (51) and were adjusted to accommodate the heterogeneous model sources. With the fitted components placed into the map, a homology model for the *E. coli* enzyme was constructed using Modeller 9.5 (52) with 10 rounds of optimization and, with CAP and DNA, were maintained as rigid bodies. The model is substantially complete with the exception of α^H residues 236–329 (α^H CTD), β residues 1–9, 223–339 (β DR1), 938–1036 (β DR2), and 1126–1179, β' residues 1–13, 1383–1407 (β' C terminus), σ^{70} subunit residues 1–112 (region 1.1), and 192–211 (acidic loop).

Model Refinement and Analysis. Geometry regularization was performed on the complete ternary complex model using phenix 1.4–3 (53). Stereochemically restrained simulated annealing refinement was performed using the Yup.scx procedure of Yup 1.090827 using default parameters (36). Model-map correlation coefficients were calculated using normal mode-based flexible fitting (NMFF) (54). UCSF Chimera (34) was used for molecular graphics analyses and imaging.

ACKNOWLEDGMENTS. We thank S. Ludtke, B. Devkota, W. Olson, D. Case, B. Carragher, and C. Potter for advice, N. Voss for assistance with image processing and generating Fig. S3, A. Leschziner for his stain protocol, Y. Ebricht for preparing DNA in a preliminary experiment, and S. Darst for providing the RNAP core map for comparison. Some of the work presented here was performed at National Resource for Automated Molecular Microscopy, supported by the National Institutes of Health (NIH) Grant RR17573 (to B. Car-

ragher and C. Potter). Reconstruction calculations were performed using the BioMaPS High Performance Computing Center, funded in part by NIH Grant RR022375 (to R. Levy). This work was supported primarily by NIH Grant

GM21589 (to C.L.L.) with additional support from NIH Grants GM41376 (to R.H.E.) and AI72766 (to R.H.E. and E.A.), and a Howard Hughes Medical Investigatorship (to R.H.E.).

- Busby S, Ebright RH (1999) Transcription activation by catabolite activator protein (CAP). *J Mol Biol* 293:199–213.
- Lawson CL, et al. (2004) Catabolite activator protein: DNA binding and transcription activation. *Curr Opin Struct Biol* 14:10–20.
- Zheng D, Constantinidou C, Hobman JL, Minchin SD (2004) Identification of the CRP regulon using in vitro and in vivo transcriptional profiling. *Nucleic Acids Res* 32:5874–5893.
- Korner H, Sofia HJ, Zumft WG (2003) Phylogeny of the bacterial superfamily of Crp-Fnr transcription regulators. *FEMS Microbiol Rev* 27:559–592.
- Schultz SC, Shields GC, Steitz TA (1991) Crystal structure of a CAP-DNA complex: The DNA is bent by 90°. *Science* 253:1001–1007.
- Parkinson G, et al. (1996) Structure of the CAP-DNA complex at 2.5 Å resolution: A complete picture of the protein-DNA interface. *J Mol Biol* 260:395–408.
- Blatter EE, Ross W, Tang H, Gourse RL, Ebright RH (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.
- Benoff B, et al. (2002) Structural basis of transcription activation: The CAP- α CTD-DNA complex. *Science* 297:1562–1566.
- Gourse RL, Ross W, Gaal T (2000) UPs and downs in bacterial transcription initiation: The role of the α subunit of RNA polymerase in promoter recognition. *Mol Microbiol* 37:687–695.
- Chen H, Tang H, Ebright RH (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.
- Ross W, Schneider DA, Paul BJ, Mertens A, Gourse RL (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.
- Naryshkin N, Revyakin A, Kim Y, Mekler V, Ebright RH (2000) Structural organization of the RNA polymerase-promoter open complex. *Cell* 101:601–611.
- Lloyd GS, Niu W, Tebbutt J, Ebright RH, Busby SJ (2002) Requirement for two copies of RNA polymerase α subunit C-terminal domain for synergistic transcription activation at complex bacterial promoters. *Genes Dev* 16:2557–2565.
- Zou C, Fujita N, Igarashi K, Ishihama A (1992) Mapping the cAMP receptor protein contact site on the alpha subunit of *Escherichia coli* RNA polymerase. *Mol Microbiol* 6:2599–2605.
- Murakami K, et al. (1997) Positioning of two α subunit carboxy-terminal domains of RNA polymerase at promoters by two transcription factors. *Proc Natl Acad Sci USA* 94:11274–11278.
- Lee DJ, Busby SJ, Lloyd GS (2003) Exploitation of a chemical nuclease to investigate the location and orientation of the *Escherichia coli* RNA polymerase α subunit C-terminal domains at simple promoters that are activated by cyclic AMP receptor protein. *J Biol Chem* 278:52944–52952.
- Mekler V, et al. (2002) Structural organization of bacterial RNA polymerase holoenzyme and the RNA polymerase-promoter open complex. *Cell* 108:599–614.
- Chiu W, Baker ML, Almo SC (2006) Structural biology of cellular machines. *Trends Cell Biol* 16:144–150.
- Haugen SP, Ross W, Gourse RL (2008) Advances in bacterial promoter recognition and its control by factors that do not bind DNA. *Nat Rev Microbiol* 6:507–519.
- Ohi M, Li Y, Cheng Y, Walz T (2004) Negative staining and image classification—powerful tools in modern electron microscopy. *Biol Proced Online* 6:23–34.
- Yu X, Hingorani MM, Patel SS, Egelman EH (1996) DNA is bound within the central hole to one or two of the six subunits of the T7 DNA helicase. *Nat Struct Biol* 3:740–743.
- Llorca O, Rivera-Calzada A, Grantham J, Willison KR (2003) Electron microscopy and 3D reconstructions reveal that human ATM kinase uses an arm-like domain to clamp around double-stranded DNA. *Oncogene* 22:3867–3874.
- Miyata T, et al. (2005) Open clamp structure in the clamp-loading complex visualized by electron microscopic image analysis. *Proc Natl Acad Sci USA* 102:13795–13800.
- Vanamee ES, Berriman J, Aggarwal AK (2007) An EM view of the FokI synaptic complex by single particle analysis. *J Mol Biol* 370:207–212.
- Hoenger A, Aebi U (1996) 3-D Reconstruction from ice-embedded and negatively stained biomacromolecular assemblies: A critical comparison. *J Struct Biol* 117:271–274.
- Mukhopadhyay J, et al. (2008) The RNA polymerase “switch region” is a target for inhibitors. *Cell* 135:295–307.
- Darst SA, et al. (2002) Conformational flexibility of bacterial RNA polymerase. *Proc Natl Acad Sci USA* 99:4296–4301.
- Artsimovitch I, Svetlov V, Murakami KS, 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.
- Iyer LM, Koonin EV, Aravind L (2004) Evolution of bacterial RNA polymerase: Implications for large-scale bacterial phylogeny, domain accretion, and horizontal gene transfer. *Gene* 335:73–88.
- Chlenov M, et al. (2005) Structure and function of lineage-specific sequence insertions in the bacterial RNA polymerase β' subunit. *J Mol Biol* 353:138–154.
- Opalka N, et al. (2000) Direct localization of a β -subunit domain on the three-dimensional structure of *Escherichia coli* RNA polymerase. *Proc Natl Acad Sci USA* 97:617–622.
- Malhotra A, Severinova E, Darst SA (1996) Crystal structure of a σ^{70} subunit fragment from *E. coli* RNA polymerase. *Cell* 87:127–136.
- Zhang G, Darst SA (1998) Structure of the *Escherichia coli* RNA polymerase α subunit amino-terminal domain. *Science* 281:262–266.
- Pettersen EF, et al. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25:1605–1612.
- Schwartz EC, et al. (2008) A full-length group 1 bacterial sigma factor adopts a compact structure incompatible with DNA binding. *Chem Biol* 15:1091–1103.
- Tan RK, Devkota B, Harvey SC (2008) YUP.SCX: Coaxing atomic models into medium resolution electron density maps. *J Struct Biol* 163:163–174.
- Murakami KS, Masuda S, Campbell EA, Muzzin O, Darst SA (2002) Structural basis of transcription initiation: An RNA polymerase holoenzyme-DNA complex. *Science* 296:1285–1290.
- Wang D, Bushnell DA, Westover KD, Kaplan CD, Kornberg RD (2006) Structural basis of transcription: Role of the trigger loop in substrate specificity and catalysis. *Cell* 127:941–954.
- Vassilyev DG, et al. (2007) Structural basis for substrate loading in bacterial RNA polymerase. *Nature* 448:163–168.
- Toulokhonov I, Zhang J, Palangat M, Landick R (2007) A central role of the RNA polymerase trigger loop in active-site rearrangement during transcriptional pausing. *Mol Cell* 27:406–419.
- Tuske S, et al. (2005) Inhibition of bacterial RNA polymerase by streptolydigin: Stabilization of a straight-bridge-helix active-center conformation. *Cell* 122:541–552.
- Cramer P, Bushnell DA, Kornberg RD (2001) Structural basis of transcription: RNA polymerase II at 2.8 Å resolution. *Science* 292:1863–1876.
- Bose D, et al. (2008) Organization of an activator-bound RNA polymerase holoenzyme. *Mol Cell* 32:337–346.
- Yang X, et al. (2009) The structure of bacterial RNA polymerase in complex with the essential transcription elongation factor NusA. *EMBO Rep* 10:997–1002.
- Suloway C, et al. (2005) Automated molecular microscopy: The new Legion system. *J Struct Biol* 151:41–60.
- Lander GC, et al. (2009) Appion: An integrated, database-driven pipeline to facilitate EM image processing. *J Struct Biol* 166:95–102.
- Mallick SP, Carragher B, Potter CS, Kriegerman DJ (2005) ACE: Automated CTF estimation. *Ultramicroscopy* 104:8–29.
- Ludtke SJ, Baldwin PR, Chiu W (1999) EMAN: Semiautomated software for high-resolution single-particle reconstructions. *J Struct Biol* 128:82–97.
- Frank J, et al. (1996) SPIDER and WEB: Processing and visualization of images in 3D electron microscopy and related fields. *J Struct Biol* 116:190–199.
- Lu XJ, Olson WK (2003) 3DNA: A software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res* 31:5108–5121.
- Larkin MA, et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947–2948.
- Eswar N, Eramian D, Webb B, Shen MY, Sali A (2008) Protein structure modeling with MODELLER. *Methods Mol Biol* 426:145–159.
- Adams PD, et al. (2002) PHENIX: Building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* 58:1948–1954.
- Tama F, Miyashita O, Brooks CL 3rd (2004) Normal mode based flexible fitting of high-resolution structure into low-resolution experimental data from cryo-EM. *J Struct Biol* 147:315–326.
- Cellai S, et al. (2007) Upstream promoter sequences and α CTD mediate stable DNA wrapping within the RNA polymerase-promoter open complex. *EMBO Rep* 8:271–278.